European Review for Medical and Pharmacological Sciences 2022; 26: 6040-6049

Research on multi-parameter fusion non-invasive blood glucose detection method based on machine learning

J.-J. LI, Z.-P. QU, Y.-W. WANG, J. GUO

School of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou, China

Abstract. – OBJECTIVE: Traditional blood glucose testing methods have several disadvantages, such as high pain and poor acquisition continuity. In response to these shortcomings, we propose a multi-parameter fusion non-invasive blood glucose detection method that combines machine learning and photoplethysmography (PPG) signal feature parameter analysis.

MATERIALS AND METHODS: This method uses the signal validity check process based on the correlation operation to test and calculate PPG data. It, then, respectively applies the bootstrap aggregation algorithm and the random forests algorithm to establish two non-invasive blood glucose detection models that comprehensively predict blood glucose data.

RESULTS: Experimental comparative analysis showed that the accuracy of the detection model based on the random forests algorithm is superior. The correlation coefficient of the obtained blood glucose prediction set is 0.972, the mean square error is 0.257, and the relative error is less than ± 20%.

CONCLUSIONS: Relative error in blood glucose prediction meets the national standards in China. Meanwhile, the results of the Clarke Error Grid Analysis indicate that the non-invasive blood glucose testing method proposed in this study meets clinical accuracy requirements.

Key Words:

Photoplethysmography, Digital signal processing, Machine learning, Non-invasive blood glucose detection.

Introduction

Daily blood glucose monitoring is an important routine that aids diabetes treatment and prevention for people with diabetes or those with similar health issues. Current clinical and individual self-testing methods for blood glucose can be divided into three categories: invasive testing, minimally invasive testing, and non-invasive testing. Invasive blood glucose testing is performed by drawing venous blood directly from the patient's body or by pricking the finger to extract a blood sample. These procedures are painful and often cause distress in certain patients. Besides, minimally invasive glucose testing methods based on subcutaneous glucose sensor microelectrode implantation[1] have the disadvantages of frequent electrode replacement, time-dependent detection accuracy, and high testing costs[2]. Non-invasive blood glucose testing overcomes the deficiencies that exist with invasive and minimally invasive methods. At present, non-invasive blood glucose detection mainly involves several technical routes, such as energy metabolism integration[3], microwave detection[4], and optical detection[5]. Of these methods, the practicality and accurate performance of optical detection methods are the most desirable[6].

Photoplethysmography (PPG) is a technique that involves the non-invasive collection of human physiological information using optical technology. It contains validated spectral data that truly reflects blood composition information[7] and can be used for the non-invasive detection of human blood glucose. Monte-Moreno et al[8] used a machine-learning algorithm to analytically fit the PPG signal collected from the fingertip site for non-invasive estimation of human blood pressure and blood glucose, with a blood glucose prediction accuracy of 87.7%. Bao et al[9] established a blood glucose concentration calculation model based on photoelectric detection technology. When combined with other human characteristic parameters, the detection results of the model achieved a correlation coefficient of 0.854. Besides, Zhang et al[10] designed a PPG signal acquisition device for the fingertip, used the wavelet threshold denoising algorithm to analyze and process PPG data, and established a blood glucose correction model based on a

Corresponding Author: Jianjun Li, D. Eng; e-mail: ljjpaper@163.com

machine learning algorithm. The blood glucose prediction set correlation coefficient of this device was 0.9152. Wang et al[11] used the photoelectric volumetric pulse wave tracing method to collect PPG data from the fingertip area and established a least squares-based detection model. This method achieved a model test set correlation coefficient of 0.976 and a root mean square error of 0.357, with an accuracy that met the requirements of clinical use.

For PPG data collection, the fingertips, neck, wrist radial artery, and other superficial artery-rich parts of the body should be selected. Currently, to obtain better quality PPG data, the fingertips are often selected as the collection site. From the perspective of practicality and portability, it is more appropriate to use the radial artery of the wrist as the collection site. However, the signal acquisition conditions in the wrist area are relatively poor. Besides, the acquired signal amplitude is weak and often contains excessive disturbance, resulting in diminished blood glucose detection accuracy. To address these problems, in this study, we investigate the signal validity check and establish a non-invasive blood glucose detection model based on valid PPG data analysis.

Materials and Methods

Acquisition of PPG Signal

The human aorta produces regular forced vibrations under the influence of the periodic beat of the heart systole and diastole. This vibration is transmitted through the blood system to the vascular endings of the body, forming a transverse wave known as a pulse wave. The pulse wave is a periodic physiological signal with a period roughly identical to the heartbeat cycle. It is also influenced by physiological factors, such as vascular wall elasticity, blood viscosity, blood flow velocity, and vascular resistance[12]. As Figure 1A indicates, during the heartbeat, the micro-arterial vessels are forced to vibrate, forming transverse waves. As a result, the blood volume of the vessels varies continually, which leads to changes in the light absorption of each blood component. When irradiated with a light source, the transmitted or reflected light intensity also varies along with changes in blood volume. This periodically changing light signal is transformed into an electrical signal, which is the PPG signal.

The PPG signal reflects the periodic variation in light intensity emitted through the human arte-

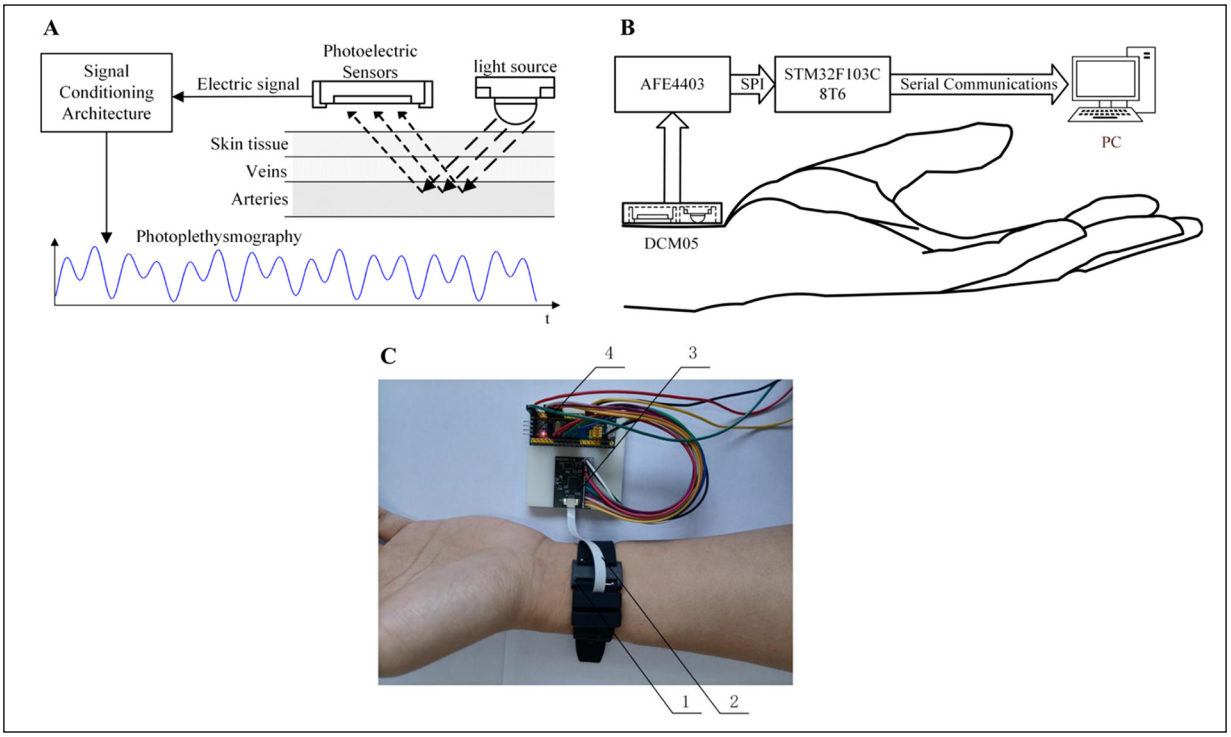

Figure 1. Acquisition of PPG. **A,** photoplethysmography; **B,** dual-wavelength reflective PPG acquisition system; **C,** schematic diagram of PPG signal acquisition (1: DCM03 acquisition sensor and protective casing, 2: signal connection line, 3: signal conditioning structure, 4: MCU).

rial vasculature. According to the Lambert-Beer law[10], the concentration c of the substance absorbing light can be measured by the incident light intensity I_0 and the outgoing light intensity I, as shown in Equation (1):

$$I = I_0 / e^{\epsilon l c} \quad (1)$$

In Equation (1), ε is the absorbance of the substance that absorbs light, and l represents the light range. Therefore, the PPG signal carries rich information on a range of blood components and can be used for the quantitative estimation of blood glucose and other blood component concentrations.

For the characteristics of skin thickness, bone, and other tissue at the wrist, a dual-wavelength reflective acquisition system was designed to complete the acquisition of PPG signals at the radial artery, as Figure 1B illustrates.

The light source and photoelectric sensor of the acquisition system in Figure 1B utilize the reflective chip sensor DCM05 (APMKorea, Daejeon, South Korea). This device integrates both 660 nm and 905 nm wavelength LED light sources and a photoelectric sensor that receives the light of corresponding wavelengths and can obtain a higher accuracy PPG acquisition signal. The front-end chip for sensor driving and signal pre-conditioning is the AFE4400 (TI, Dallas, TX, USA) integrated analog front-end. It is commonly used for heart rate monitoring, pulse blood oxygen calculation, and industrial light measurement applications to meet system usage and accuracy requirements. Besides, the system uses the STM32F103ZE chip (ST, Geneva, Canton of Geneva, Switzerland) as the main control unit (MCU). The MCU communicates with the host computer through the serial protocol to complete further processing of data and feedback on blood glucose values. Considering that the effective frequency of the PPG signal is between 0.7~3 HZ[11], the system digital sampling frequency is selected as 50 HZ. This setting satisfies the Nyquist theorem and can ensure the acquisition accuracy of the signal. Also, by considering the fixation of the sensor and the portability of the device, the acquisition system uses the wristband fixation method for the acquisition process of the PPG signal, as illustrated in Figure 1C.

The Principle of Signal Validity Check

Figure 2A introduces possible external disturbances, such as pulse interference, signal loss, and motion artifacts[13] in the initial PPG signal obtained from the acquisition system. These disturbances are often caused by subject jitter, ambient light interference, or sensor slippage. Since these disturbances that may exist in the acquired signal and cannot always be effectively filtered out during the filtering process, blood glucose detection accuracy can be affected. To address this problem, we propose a PPG acquisition signal validity check method based on the correlation operation, which can accurately identify and reject signal segments with excessive disturbance to ensure the validity of the collected signals.

The validity check of the PPG signal is implemented based on the correlation operation[14]. The correlation operation is similar to the convolution operation, which is an operation that combines two signal sequences to obtain a third signal. It can be used to detect the presence or absence of a target signal in the received signal.

The convolution operation process can be represented by Equation (2):

$$y^c = x^c * h^c \quad (2)$$

In Equation (2), x^c denotes the discrete input sequence, h^c signifies the convolution kernel, and y^c is the output sequence. The convolution operation determines y^c by reversing the order of the sample points in h^c, then multiplying and adding the sample points in x^c. If the length of x^c is M and the length of h^c is N, the length of y^c can be expressed as N+M-1. The computational process of the correlation operation is similar to that of the convolution operation, and the resulting third signal is called the cross-correlation signal of the two input signals. The output-side algorithm for the correlation operation can be expressed by Equation (3):

$$y^r(i) = \sum_{j=0}^{F-1} h^r(j) x^r(i+j) \quad (3)$$

In Equation (3), y^r represents the cross-correlation signal, x^r denotes the input signal, h^r stands for the target signal, and F represents the signal length of h^r. Also, i and j are the sample numbers of the signal in Equation (3). Equation (3) indicates that the difference between the two operations is that the correlation operation does not reverse the order of the samples in h^r, but directly multiplies and adds the samples in x^r to obtain y^r, where the signal length of y^r is equal to x^r.

According to the correlation operation, the amplitude of each sample point of y^r depends on the similarity between x^r and h^r at the current sample point, so that each moment of h^r appearing in x^r corresponds to a symmetrical wave peak in y^r. Ideally, the value of the peak is expressed using Equation (4):

$$y^r(i)_{\max} = \sum_{j=o}^{F-1} h^r(j)^2 \tag{4}$$

Therefore, by counting the number of such peaks, the quantity of the target signal h^r contained in the input signal x^r can be obtained. According to this principle, the validity of PPG acquisition signal can be checked.

The main purpose of the signal validity check is to screen and reject the acquired signal segments with excessive disturbance. Thus, the target signal in the PPG signal acquisition validity check process is selected as the standard PPG signal. Furthermore, considering the accuracy of signal detection, the signal length of the target signal should not be too large. It is most appropriate for the signal length to only contain one cycle of the standard PPG signal.

The standard PPG signal of one cycle consists of three waveforms: the main wave, dichotomous wave, and dichotomous pre-wave, which can be simulated by three Gaussian functions. Thus, the superposition of the three corresponding Gaussian functions constitutes a period of the simulated standard PPG signal[15], so the design of the target signal h^r_{PPG} in this study can be expressed using Equation (5):

$$\begin{cases} h^r_{PPG}(t) = PPG_{sim}(t) \\ PPG_{sim}(t) = \sum_{i=1}^{3} A_i \exp[\frac{-(t-P_i)^2}{B_i}] \end{cases} \tag{5}$$

In Equation (5), t stands for the sample number of h^r_{PPG} and PPG_{sim} represents a period of the analog standard PPG signal sequence. Besides, the parameters A_i, P_i, and B_i determine the height, position of the peak, and width of the Gaussian waveform, respectively. To facilitate effective wave peak screening, the amplitude of PPG_{sim} is taken as the maximum value of the PPG acquisition signal amplitude in the system. Also, to simulate the actual sampling frequency of the system, the data sample interval t_c in h^r_{PPG} is represented by Equation (6):

$$t_c = 1 / f_c \tag{6}$$

In Equation (6), f_c signifies the system sampling frequency, which is 50 HZ in this study. Additionally, to allow the signal period of PPG_{sim} to match the period of the PPG signal contained in the current acquisition signal, the signal period of PPG_{sim} is expressed using Equation (7):

$$len_h = len_s = f_c / f_s \tag{7}$$

In Equation (7), len_h denotes the period length of PPG_{sim}, len_s represents the period length of the current actual PPG acquisition signal, and f_s is the main frequency of the actual PPG acquisition signal. The corresponding signal frequency spectrum can be obtained using the fast Fourier transform of the original acquired signal with the baseline drift filtered out. Also, the frequency corresponding to the maximum value of the signal spectrum between 0.7 HZ and 3 HZ effective frequency is taken as the main frequency f_s of the effective PPG signal contained in the current acquired signal. Combined with the analysis of the acquired signal waveform, the values of each parameter[16] in Equation (5) can be determined and are shown in Table I.

By combining Equations (5), (6), and (7) and the parameters in Table I, the corresponding target signal required by the current acquisition signal validity process can be obtained.

Figure 2B shows the linear amplitude spectrum obtained by the fast Fourier transform after filtering out the baseline drift of the analog interference signal shown in Figure 2A. As Figure 2B illustrates, the frequency corresponding to the maximum spectral amplitude of the analog

Table I. Value of parameters in h^r_{PPG}

Parameter	Value (mV)	Parameter	Value	Parameter	Value
A_1	1.5	B_1	$30/f_s$	P_1	$16/f_s$
A_2	1.125	B_2	$30/f_s$	P_2	$30/f_s$
A_3	0.375	B_3	$30/f_s$	P_3	$40/f_s$

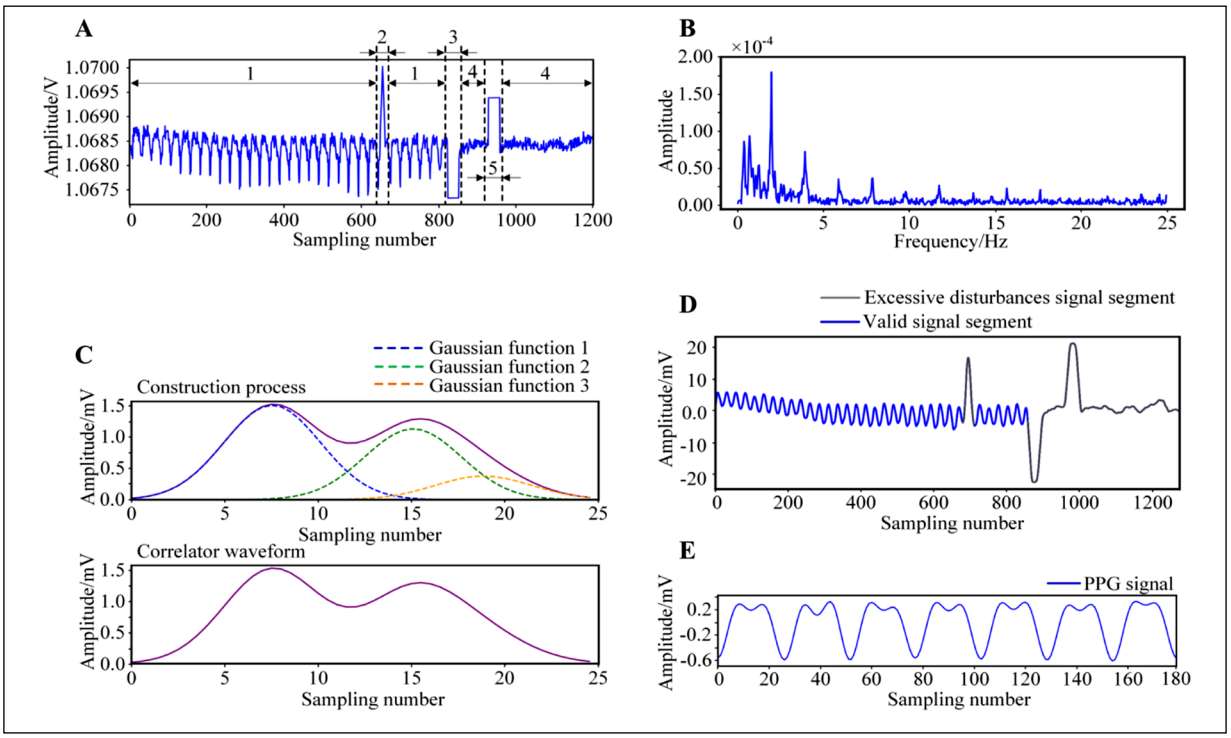

Figure 2. The principle of signal validity check and the filtered signal. **A,** analog interference signal (1: valid signal segment, 2: motion artifact, 3: low pulse interference, 4: signal loss segment, 5: high pulse interference); **B,** linear amplitude spectrum; **C,** target signal; **D,** cross-correlation signal; **E,** filtered PPG acquisition signal at the radial artery.

interference signal between 0.7 HZ and 3 HZ is 2 HZ, so the f_s of the effective PPG signal contained in the analog interference signal is 2 HZ. Meanwhile, according to Equation (7), the corresponding target signal length is 25 sample points. Figure 2C displays the construction process of the corresponding target signal, which can be obtained after combining Equations (5) and (6) and the parameters in Table I.

Figure 2D presents the cross-correlation signal obtained by the correlation operation between the analog interference signal after baseline drift is filtered and the corresponding target signal. Corresponding effective peaks occur in the cross-correlation signal at the position where the target signal appears in the acquired signal. However, the cross-correlation signal corresponding to the excessive disturbance signal segment does not contain any effective peaks, meaning it is an irregular waveform signal.

PPG Acquisition Signal Validity Check Process and Filtering

The validity of the PPG acquisition signal is checked in terms of data segments, which are intercepted from the initial position of the original acquisition data X^r. The interception process of data segments can be expressed by Equation (8):

$$x_n = X^r\,[n \times 2000\colon ((n+1) \times 2000 - 1)] \tag{8}$$

In Equation (8), x_n represents data segments, where the subscript n is the current data segment number. n starts at 0 and x_n length is 2,000 sample points containing 40 seconds of collected data. The validity check process for x_n is shown in Figure 3.

The process shown in Figure 3 starts with high pass filtering of x_n to eliminate baseline drift and avoid interference with peak detection from the DC component of the signal. In Figure 3, x^L_n is the high pass filtered signal. The corresponding spectrum $x^L_{n\_FFT}$ is then obtained by performing the fast Fourier transform on x^L_n. The f_s of the current signal segment can be obtained from $x^L_{n\_FFT}$, and the corresponding target signal h^n_{PPG} of the current signal segment is constructed according to f_s. After obtaining

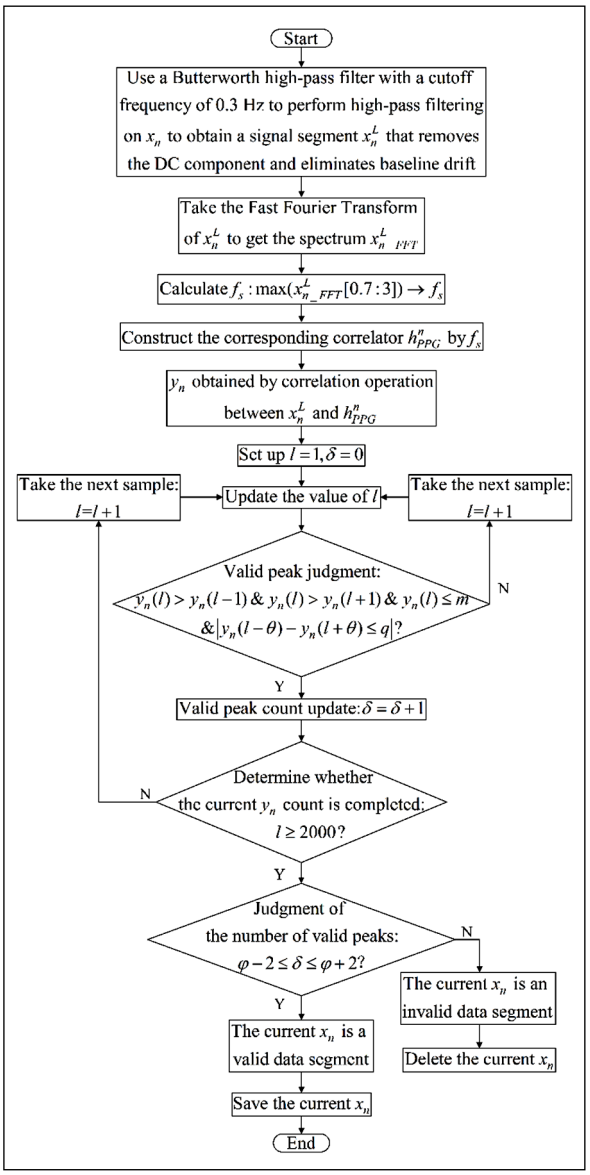

Figure 3. Signal validity check process.

$h^n{}_{PPG}$, $x^L{}_n$ is correlated with $h^n{}_{PPG}$ to obtain the cross-correlation signal y_n. Besides, in Figure 3, l represents the sample number of the data in y_n. The peak threshold of the valid waveform in y_n is denoted as m in the check process, and the value of m is calculated using Equation (9):

$$m = \sum_{j=0}^{len_h} h_{PPG}^r[j]^2 \tag{9}$$

Meanwhile, due to the difference between the amplitudes of $h^n{}_{PPG}$ and the PPG signal in the collected signal, the valid wave peaks in the cross-correlation signal cannot satisfy absolute symmetry. Consequently, in this study, the wave peaks satisfying Equation (10) in y_n are approximated to be symmetrical.

$$\begin{aligned} &|y(l-\theta)-y(l+\theta)|\le q, \\ &\theta\in\left\{N^*\,|[1,((f_c/f_s)-1)/2]\right\} \end{aligned} \tag{10}$$

In Equation (10), θ denotes the sequence number of the signal sample point and q signifies the empirical threshold, which is calculated by the waveform analysis of the acquired signal and the cross-correlation signal. In this study, the value of q is 0.0016 mV. Therefore, if the sample $y_n(l)$ in y_n is a valid peak point, $y_n(l)$ should satisfy Equation (11):

$$\begin{cases} y_n(l)>y_n(l-1) \\ y_n(l)>y_n(l+1) \\ |y_n(l-\theta)-y_n(l+\theta)|\le q \\ y_n(l)\le m \end{cases} \tag{11}$$

In Figure 3, δ is a valid peak count parameter, and the value of δ is updated every time a valid peak is detected in y_n. Besides, φ represents the number of valid peaks that should theoretically be included in y_n if the current x_n is a valid signal segment. The value of φ can be determined using Equation (12):

$$\varphi = 2000 / (f_c / f_s) \tag{12}$$

Therefore, considering the loss of signal period caused by interception of the data segment, when the current data segment x_n with a length of 2,000 samples is a valid signal segment, the current δ should satisfy Equation (13):

$$\varphi - 2 \le \delta \le \varphi + 2 \tag{13}$$

After completing the count of valid peaks in y_n, the process checks whether δ satisfies Equation (13). If it is satisfied, the corresponding x_n is determined to be a valid data segment. Otherwise, x_n is determined to be invalid and cannot be used for blood glucose detection so it is eliminated, and the next data segment is intercepted for checking.

The signal x_n passing the validity check contains some slight disturbances such as power frequency interference, EMG interference, and random noise. Signal filtering can be accom-

plished using an 8th-order Butterworth filter. Figure 2E shows a partial waveform of the PPG signal collected from the radial artery after filtering. It can be seen that the filtered signal has a complete shape, no evident noise interference, and clear waveform characteristics. Therefore, it can be used to establish a non-invasive blood glucose detection model.

Non-Invasive Blood Glucose Detection Feature Parameter Extraction and Detection Model Establishment

In this study, the data segments x_n, which passed the validity check and filtering process, were used as samples for the extraction of feature parameters. When extracting parameters, the data segment is initially divided into multiple data windows. The data windows are denoted as w_τ, where τ represents the data window sequence number. Each w_τ is 200 sample points in length and contains 4-6 pulse beats in about 5 seconds, which is smaller than the human respiratory cycle and can reduce the interference of respiratory action and other physiological activities[17]. The adjacent windows are designed to have a 50% data overlap, so a 2,000-point data set can be divided into 19 windows. The w_τ is used as the minimum unit for signal feature parameter extraction calculations.

The PPG signal contains rich amounts of information. Therefore, for the model in this study, we selected an assortment of parameters that are more typical and have a higher correlation with blood glucose as the feature parameters.

The spectroscopic entropy H is a common characteristic value for quantitative spectral analysis and represents the entropy value of a pure substance[18]. To calculate H in terms of w_τ, the frequency spectrum W_τ is first obtained using the fast Fourier transform of w_τ. Next, P^τ_W is obtained by normalizing W_τ so the spectroscopic entropy $H^\tau_w(k)$ on w_τ can be expressed by Equation (14)[10]:

$$H^\tau_w(k) = P^\tau_W(k) \log(P^\tau_W(k)) \tag{14}$$

The H value of the data segment can be calculated by superimposing $H^\tau_w(k)$. Besides, the mean H^μ, variance H^σ, quadratic spacing H^{iqr}, and skewness H^{skew} of H on the data segment are used as the components of the model input matrix.

In diabetic patients, blood pressure and blood glucose have a direct correlation. Moreover, the spectral energy logarithm feature E of the PPG signal is closely related to the respiratory rate and can indirectly reflect the blood pressure level. Therefore, E is used as a feature parameter of the model. E_τ, the spectral energy logarithm feature on w_τ, can be expressed using Equation (15)[19]:

$$\log E_\tau = \log\left(\sum_{k=0}^{199} w_\tau(k)^2\right) \tag{15}$$

In the data segment, E is obtained by the superposition of E_τ, and the statistical values E^σ and E^{iqr} of E can be used as components of the model input matrix.

The pulse wave transmission time T is related to the pulse wave transmission velocity, and there is also a correlation between the pulse wave transmission velocity and arterial blood pressure[20]. Additionally, because the blood pressure of diabetic patients is closely related to blood glucose, T can also be used as a feature parameter of the model. The pulse wave transmission time on w_τ can be obtained by calculating the length of each cycle of the PPG signal in w_τ. The mean value T^μ and the variance T^σ of the pulse wave transmission time in each w_τ of the data segment are used in the composition of the model input.

As a measure of the oxygen content in the blood, blood oxygen saturation (SpO_2) is an important parameter of the human respiratory system. When abnormal blood glucose levels affect the respiratory function, it leads to a lower SpO_2 value[9], so SpO_2 can be used as a feature parameter of the model. The calculation of SpO_2 of the tested sample can be expressed by Equation (16):

$$SpO_2\, k_1 + k_2(I_R / I_{IR}) + k_3 (I_R / I_{IR})^2 \tag{16}$$

In Equation (16), I_R and I_{IR} respectively represent the reflected light intensity of red light and near-infrared light collected by the system. They can be obtained by reading the LEDVA and ALED-VAL registers of AFE4400, where k_1, k_2, and k_3 represent empirical scaling constants. Besides, SpO_2 is the last component of the model input.

By combining the above parameters, two non-invasive blood glucose detection models were established based on either the bootstrap

aggregation algorithm or the random forests algorithm. The model input matrix Q can be expressed by Equation (17):

$$Q=[H_i^{\mu},H_i^{\sigma},H_i^{iqr},H_i^{skew},E_i^{\sigma},E_i^{iqr},T_i^{\mu},T_i^{\sigma},SpO_{2i}] \quad (17)$$

In Equation (17), i denotes the sample number.

Results

Blood Glucose Prediction Results

To verify the correctness and accuracy of the signal validity check process and the blood glucose detection method, we enlisted 40 volunteers aged 20-60 years old for data collection and testing. During data collection, the volunteers maintained even breathing, kept calm, relaxed their arms, and opened their palms naturally. The PPG data was collected from the radial artery of the volunteers' wrists, and an invasive blood glucose meter was used also to detect the true blood glucose value. Finally, 107 sets of data that passed the validity check were obtained. Then, the collected data was filtered, and the feature parameters extracted. Model training was performed based on the bootstrap aggregation algorithm and the random forests algorithm, respectively. After training, we obtained two non-invasive blood glucose detection models with specific accuracy, which were used to determine blood glucose prediction values. The curves of the blood glucose prediction values of the two algorithm models and the true values are presented in Figure 4A and 4B.

Error Analysis of Blood Glucose Prediction

Figure 4A and 4B indicates that the blood glucose prediction models based on the two algorithms have better prediction accuracy in the middle range of blood glucose values. The relative error curves of the prediction results of the two models are displayed in Figure 4C, which shows that the relative error of the random forests algorithm model is always less than ±20%. These results are more accurate than the bootstrap aggregation algorithm model.

The correlation coefficient and mean square error (MSE) of the prediction results of the two algorithm models are shown in Table II. The correlation coefficient and MSE of the random

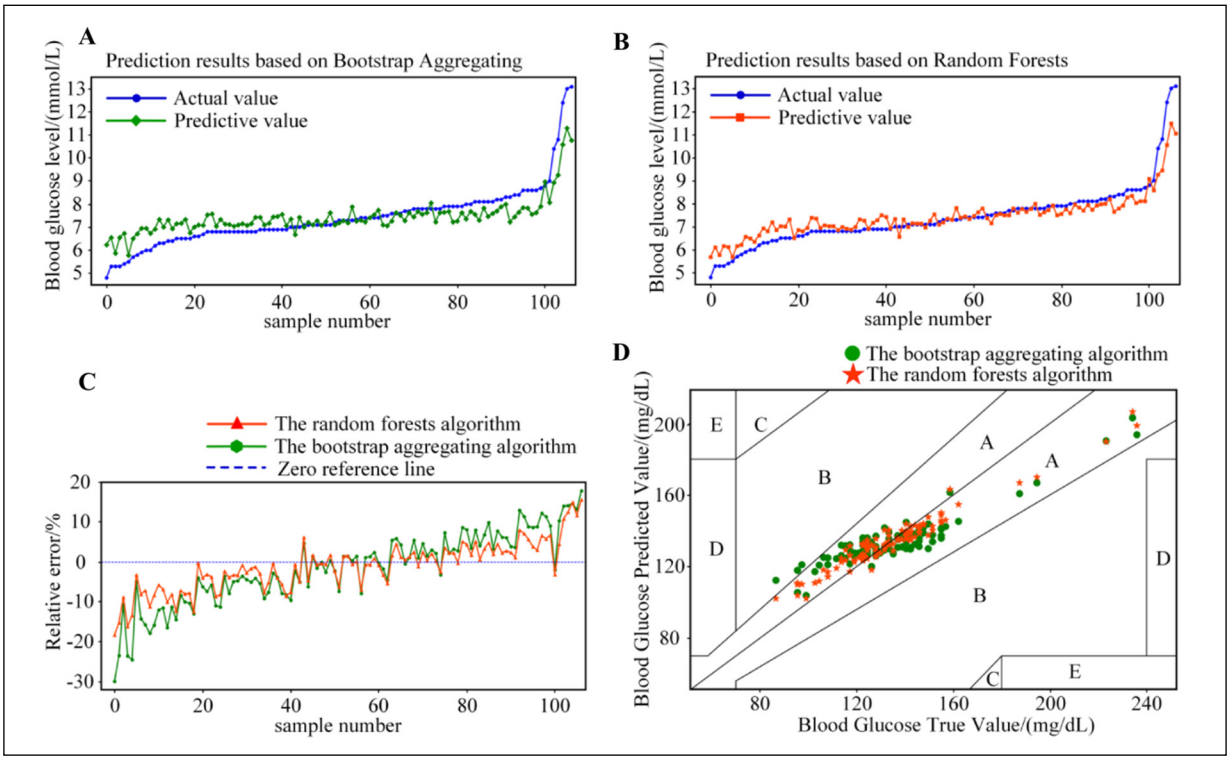

Figure 4. Blood glucose prediction results analysis. **A,** blood glucose prediction results based on the bootstrap aggregation algorithm model; **B,** blood glucose prediction results based on the random forests algorithm; **C,** relative errors of the predicted values of the two models; **D,** Clarke Error Grid Analysis of predicted blood glucose values.

Table II. Correlation coefficient and MSE of algorithm model.

Model algorithm	Correlation coefficient	MSE
Random forests	0.972	0.257
Bootstrap aggregation	0.924	0.474

forests model results are better than those of the bootstrap aggregation model.

To further compare the accuracy of the two models, we introduced the Clarke Error Grid Analysis[21] to analyze the blood glucose prediction values of the algorithms. The Clarke Error Grid Analysis was developed in 1987 and is an analytical tool specially used for evaluating the accuracy of blood glucose detection values.

The Clarke Error Grid Analysis establishes a two-dimensional scatter plot using the true blood glucose value and the blood glucose detection value as the horizontal and vertical coordinates, respectively. This coordinate system is divided into five distinct regions: A, B, C, D, and E. Scattered points in area A indicate that the absolute percentage error between the blood glucose detection value and the true value does not exceed 20%, thereby meeting clinical requirements. The points in zone B suggest that there is an obvious deviation between the blood glucose detection value and the real value. It does not meet clinical standards but does not lead to misdiagnosis. In region C, the error is too large and there is no reference value. Zone D is a dangerous error area, and the data in this area may lead to misdiagnosis. Finally, E is the complete error area, where the blood glucose detection value does not correlate at all with the true value. The Clarke Error Grid Analysis results of the predicted values of the two algorithm models are shown in Figure 4D.

Figure 4D indicates that for the random forest's algorithm model, 99% of the sample points in the blood glucose prediction value fall within area A. In contrast, 97% of the sample points for the bootstrap aggregation algorithm are in area A. Therefore, based on the Clarke Error Grid Analysis results, the accuracy of the prediction model of the random forests algorithm is better than that of the bootstrap aggregation model.

Discussion

Analysis of the experimental results shows that there is a correlation between the feature parameters selected in this study and blood glucose value. Besides, the signal validity check process and blood glucose detection method are also correct. The relative error, correlation coefficient, MSE, and Clarke Error Grid Analysis results of the blood glucose prediction values all show that the detection model based on the random forests algorithm has superior accuracy. Therefore, it is more suitable for blood glucose detection.

Conclusions

The main research focus of this paper is the validity check of PPG signals and the non-invasive detection of blood glucose based on machine learning.

In this study, a reflective PPG signal acquisition device was first designed based on photoplethysmography. Using the radial artery of the wrist as the acquisition position, a primary PPG signal was obtained. Meanwhile, considering the influence of possible excessive waveform disturbance on the accuracy of blood glucose detection, we proposed a PPG signal validity check process based on correlation operations to screen out disturbance from the PPG signal. After completing the validity check and filtering the PPG signal, we established two non-invasive blood glucose detection models based on the bootstrap aggregation algorithm and the random forests algorithm, respectively. Then, we selected certain variables that were highly correlated with blood glucose as the input vector of the blood glucose detection model. These variables included spectral entropy features, spectral logarithmic features, pulse wave transmission time, and SpO_2. An analysis of experimental results shows that the non-invasive blood glucose detection model based on the random forests algorithm has higher accuracy. More specifically, the correlation coefficient of the obtained blood glucose prediction set is 0.972, the MSE is 0.257, the relative error is less than ±20%, and 99% of the sample points in the Clarke Error Grid Analysis are in zone A. The blood glucose prediction results meet the relevant national standards[22] and also the requirements for clinical accuracy.

The PPG signal has strong variability and contains rich information. Therefore, the next step of this research is to increase the input dimension of the blood glucose detection model and select more PPG feature parameters as the inputs of the model. Additionally, the amount of data collected for people with various physical conditions will be increased, and the relevance and overall accuracy of the model will be enhanced.

Conflict of Interest
The Authors declare that they have no conflict of interests.

Ethics Approval and Informed Consent
This article does not contain any studies with human participants or animals performed by any of the authors.

Availability of Data and Materials
Data can be provided upon reasonable request.

Authors' Contribution
Jianjun Li and Zaipeng Qu participated in processes of theoretical research, experimental design, data collection and analysis, writing and quality evaluation. Yiwen Wang and Jing Guo participated in processes of data collection and quality evaluation.

References

1) Liu YX, Yao YF, Huang B, Zhong M. A MEMS Sensor for Blood Glucose Monitoring Based on Tissue Fluid Ultrafiltration Sampling. Nanotechnol Precis Eng 2013; 11: 202-210.
2) Ropmay GD, Rangababu P, Akhtar J, Kumar S, Rathore PK. A MEMS based Blood Glucose Measurement Sensor Using. AIP Conference Proceedings 2020; 2294: 1-6.
3) Zhang Y, Chen ZC, Zhu JM, Xu B. An Improved Method for Noninvasive Detection of Blood Glucose of Conservation of Energy Metabolism. Chin J Sens Actuators 2016; 29: 808-812.
4) Baghelani M, Abbasi Z, Daneshmand M, Light PE. Non-invasive continuous-time glucose monitoring system using a chipless printable sensor based on split ring microwave resonators. Sci. Rep 2020; 10: 1-15.
5) Cheng JX, Ji Z, Li MZ, Dai J. Study of a noninvasive blood glucose detection model using the near-infrared light based on SA-NARX. Biomed Signal Proces 2020; 56: 1-10.
6) Heise HM, Marbach R, Koschinsky T, Gries FA. Noninvasive blood glucose sensors based on near-infrared spectroscopy. Artif Organs 2010; 18: 439-447.
7) Liu J, Cai ZJ, Zhang ZY, Sun D, Wong W. Evaluation on noninvasive blood components measurement based on photolepthysmography. Opt Precis Eng 2016; 24: 1264-1271.
8) Monte-Moreno E. Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. Artif Intell Med 2011; 53: 127-138.
9) Bao Y, Shan XZ, Wang GX, Hong RJ, Zhong DW, Gao XM. A research on data processing method for non-invasive blood glucose based on detection information fusion. Optical Instruments 2017; 39: 1-7.
10) Zhang BX. Research on the Intelligent Noninvasive Blood Glucose Detection System. Lanzhou University of Technology 2020.
11) Wang XF, Tang HT, Zhang XY. Non-invasive Blood Glucose Detection System Based on Near-infrared Light. Res Explor Lab 2021; 40: 87-90.
12) Qiao AK, Wu SG. Theories of Pulse Wave in Arteries. J Biomed Eng 2000; 106: 95-100.
13) Chou, YX, Zhang AH, Wang P, Gu J. Pulse rate variability estimation method based on sliding window iterative DFT and Hilbert transform. J Med Biol Eng 2014; 34: 347-355.
14) Smith-Steven W. Digital Signal Processing: A Practical Guide for Engineers and Scientists [M]. Beijing: People's Posts and Telecommunications Publishing House, 2010: 106-110.
15) Martin-Martinez D, Casaseca-de-la-Higuera P, Martin-Fernandez M, Alberola-Lopez C. Stochastic Modeling of the PPG Signal: A Synthesis-by-Analysis Approach with Applications. IEEE Trans Biomed Eng 2013; 60: 2432-2441.
16) Qian WL, Xu LY, Chen FY, Zheng RL. Acquiring Characteristics of Pulse Wave by Gauss Function Separation. Chin J Biomed Eng 1994; 15: 1-7.
17) Ma S. NBGM Model and Algorithm Research Based on PPG. Graduate School of Chinese Academy of Sciences (Shenyang Institute of Computing Technology) 2016.
18) Sun MX, Chen NG. Non-invasive measurement of blood glucose level by time-resolved transmission spectroscopy: A feasibility study. Opt Commun 2012; 285: 1608-1612.
19) Chuah ZM, Paramesran R, Thambiratnam K, Poh SC. A two-level partial least squares system for non-invasive blood glucose concentration prediction. Chemometr Intell Lab 2010; 104: 347-351.
20) Wang B, Yang Y, Xiang J. A Noninvasive Method for Radial Pulse-Wave Velocity and the Determinants of Pulse-Wave Velocity. J Biomed Eng 2000; 17: 179-182.
21) Clarke WL, D Cox, Gonderfrederick LA, Carter W, Pohl SL. et al. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care 1987; 23: 622-628.
22) China State Food and Drug Administration. In vitro diagnostic test systems-General technical requirements for blood-glucose monitoring systems for self-testing: GB/T 19634—2005 [S]. Beijing: Beijing Medical Device Inspection Institute, 2005.