Scientific Data

https://doi.org/10.1038/s41597-026-06945-6

Article in Press

A Four-Wavelength Photoplethysmography dataset for non-invasive hemoglobin assessment

Received: 30 June 2025

Accepted: 20 February 2026

Cite this article as: Chen, L., Li, S., Liu, L. *et al.* A Four-Wavelength Photoplethysmography dataset for non-invasive hemoglobin assessment. *Sci Data* (2026). https://doi.org/10.1038/s41597-026-06945-6

Liangqiu Chen, Shiyong Li, Lei Liu, Guanfeng Huang, Yan Zhuang, Zhencheng Chen, Yongbo Liang & Mohamed Elgendi

A Four-Wavelength Photoplethysmography dataset for non-invasive hemoglobin assessment

Liangqiu Chen[1,#], Shiyong Li[2,#], Lei Liu[1], Guanfeng Huang[3], Yan Zhuang[4], Zhencheng Chen[1,5], Yongbo Liang[1,5,6,*] & Mohamed Elgendi[7,8,*]

[1] School of Life and Environmental Sciences, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China

[2] School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China

[3] The Hospital of Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China

[4] National Supercomputing Center in Xi'an, Shaanxi, 710100, China

[5] Guangxi Key Laboratory of Metabolic Reprogramming and Intelligent Medical Engineering for Chronic Diseases, Guilin,Guangxi, 541004, China

[6] Guangxi Center for Applied Mathematics (Guilin University of Electronic Science and Technology)

[7] Department of Biomedical Engineering and Biotechnology, Khalifa University of Science and Technology, Abu Dhabi, UAE

[8] Center for Biotechnology, Khalifa University of Science and Technology, Abu Dhabi, UAE

# Liangqiu Chen and Shiyong Li contributed equally to the article.

* Corresponding authors: mohamed.elgendi@ku.ac.ae (M. Elgendi), liangyongbo@guet.edu.cn (Y. Liang)

Abstract

Hemoglobin (Hb) concentration is a fundamental physiological marker widely used in the diagnosis of anemia and the assessment of cardiovascular health. Although invasive blood testing provides high accuracy, its reliance on laboratory infrastructure limits scalability and real-time applicability. Here, we present Hb-PPG, a four-wavelength photoplethysmography (PPG) dataset designed to support research on non-invasive hemoglobin assessment and cardiovascular monitoring. The dataset comprises 1008 PPG signal segments acquired at 660, 730, 850, and 940 nm from 252 adult subjects, alongside reference measurements of hemoglobin, fasting blood glucose, and brachial artery systolic and diastolic blood pressure. Hb-PPG enables systematic investigation of wavelength-dependent PPG signal characteristics and their relationships with hematological and hemodynamic parameters. By providing high-quality, multi-wavelength optical signals with clinically grounded reference data, this dataset facilitates the development, validation, and benchmarking of non-invasive approaches for hemoglobin estimation and related vascular health applications.

Background and Summary

Hemoglobin (Hb) is the major protein in blood erythrocytes and is responsible for transporting oxygen from the lungs to tissues and organs throughout the body, and is a key molecule in maintaining the balance of oxygen metabolism in the body[1]. The total Hb (tHb) concentration in the blood is an important clinical indicator, and abnormalities in its level may indicate the presence of disease or pathology in the body[2]. Specifically, high Hb levels may lead to increased blood viscosity, increasing the risk of thrombosis and cardiovascular disease, while low Hb levels may lead to anemia, causing a range of symptoms such as fatigue, malaise, and cognitive decline, with a particularly pronounced impact on susceptible populations such as women, children, and the elderly[3,4,5]. Thus, accurate detection of Hb in blood is important for early prevention and diagnosis of diseases.

Currently, the gold standard for Hb assessment is invasive venous blood sampling combined with laboratory analysis[6]. This method relies on specialized technicians and complex equipment, which is costly, time-consuming, and is likely to cause discomfort to patients. In view of limitations of traditional Hb detection methods, more and more studies have focused on the development and application of non-invasive detection technologies in clinical fields[7,8,9]. Among these, photoplethysmography (PPG) technology stands out as a research hotspot in the field of non-invasive Hb detection due to its advantages, such as easy operation, low cost, stable performance, and strong adaptability[10,11,12]. PPG is a non-invasive technique based on optical principles for detecting changes in peripheral circulating blood volume. This principle is based on the Beer-Lambert law, in which a light beam with a specific wavelength is directed at the surface of skin rich in blood vessels, which is transmitted or reflected and subsequently received by a photoreceiver.

In this process, the light signal is attenuated when absorbed by skin, muscle and blood, and the degree of attenuation is proportional to the concentration of the substance in blood and light travel length, finally results in PPG signal[13]. PPG signal contains a wealth of physiological and pathological information, and it can be used to detect hemoglobin, oxygen saturation[14], respiratory rate, blood glucose[15], blood pressure[16], arterial stiffness[17], and other physiological information.

Through analyzing the optical properties of blood, it has been found the absorption of a particular wavelength beam by blood is mainly determined by the oxygen saturation of hemoglobin[18]. In the spectral range of 600 nm to 1000 nm, hemoglobin exhibits unique absorbance properties at different wavelengths, reflecting the changes in the concentration of oxyhemoglobin and deoxyhemoglobin[19]. Regard to different light wavelength, one single wavelength can only provide limited information and usually only measure the concentration of oxyhemoglobin or deoxyhemoglobin, making it difficult to distinguish them accurately. In contrast, multi-wavelength technique is able to capture more hemodynamic changes by using different light wavelengths penetrating tissues at different depths, thus reflecting more comprehensively changes in blood components[20]. Based on this, we acquired multi-wavelength PPG (MW-PPG) signal to disclose hemodynamic changes of body blood and tissues in this study.

The Hb-PPG dataset consists four-wavelength (660 nm, 730 nm, 850 nm and 940 nm) PPG signal collected from 252 adult subjects, aged 21–90 years, with a median age of 47 years. Females accounted for 56.7%. The dataset covers groups with different Hb and blood glucose concentration levels, and also records brachial artery blood pressure values. It provides opportunities for exploring relations between PPG signal and Hb, and conducting joint analysis of multiple physiological parameters. In the following section, we will describe the dataset building process, dataset detail records, dataset usage descriptions and explain its potential application scenarios.

Methods

● Experimental design

We established a non-invasive hemoglobin detection system, which consists of two parts: hardware device and a host computer software. The hardware sensor used includes a multi-mode sensor front-end chip (ADPD4100, Analog Devices Inc.), a contact pressure sensor (FSR400, Interlink Electronics), a PPG sensor (DCM08, APM Korea), and a microcontroller unit (STM32F411, Core ARM Cortex-M4 100 MHz). These hardware sensors cooperate with each other to complete the process from signal acquisition to data transmission. Specifically, the PPG sensor is responsible for transmitting multi-wavelength optical signals and receiving the reflected optical signals to realize PPG signal

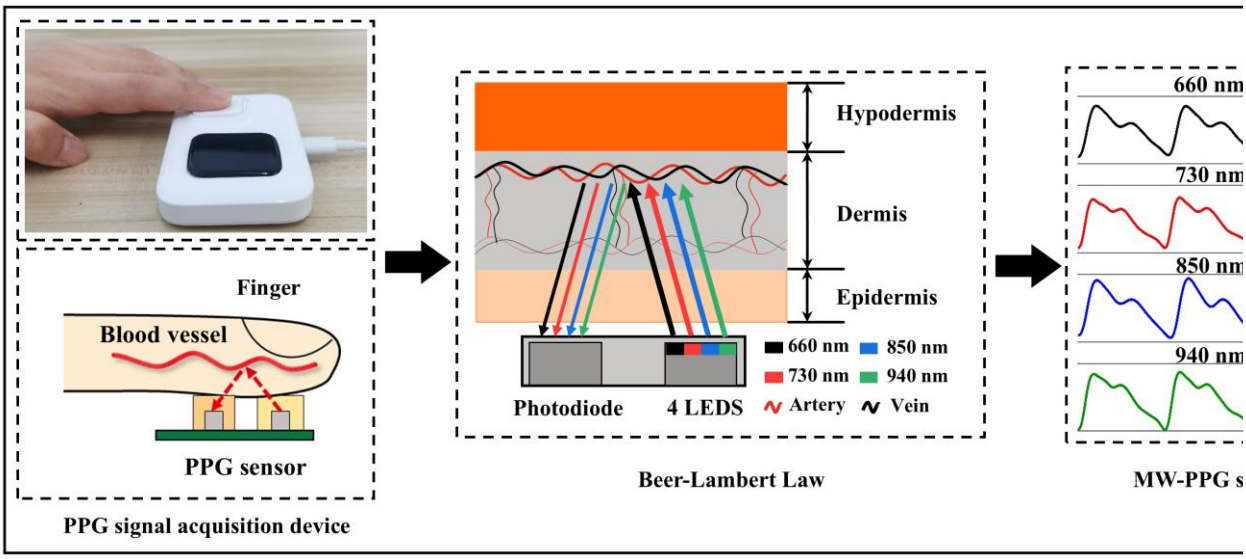

Figure 1. Illustration of the data acquisition device and working principles of PPG sensor. LEDS: light emitting diode(s). MW-PPG: multi-wavelength PPG.

recording. The wavelengths emitted include 660 nm (red light), 730 nm (near infrared light), 850 nm (infrared light) and 940 nm (infrared light), which covered the wavebands sensitive to the absorbance characteristics of hemoglobin[21]. The front-end chip is responsible for initial processing of the original MW-PPG signal, it processes four optoelectronic signals, conducts the current-to-voltage conversion, filters and amplifies, and converts the analog signal into the corresponding digital signal for transmission to the master control chip. The contact pressure sensor captures pressure changes between the PPG sensor and fingertip, ensures the sensor maintains normal contact with the subject skin, and transmits the pressure signal to the microcontroller unit for subsequent processing. The microcontroller unit, as the core control unit of the system, assumes many functions. It interacts with the front-end chip via serial peripheral interface (SPI) communication to acquire the MW-PPG digital signal in real time. Meanwhile, it utilizes analog-to-digital (ADC) module to convert the analog pressure signal acquired by the pressure sensor into a digital signal. Finally, the microcontroller unit transmits these data to the laptop via serial communication for further analysis and processing by the host computer software.

The design of host software is based on the computer interface, which can communicate with the MW-PPG acquisition device through the serial port or Wi-Fi, acquire the MW-PPG signal in real time and store its data in the local SQLite database. Before data acquisition, users need to enter the physiological information of subjects and prompt the subjects to sit still and prepare for acquisition; During the acquisition process, the software interface is eligible to display the remaining time of acquisition and contact pressure prompts in real time, which will help subjects better calibrate the applied force to PPG sensor and ensure data acquisition was carried out smoothly. Meanwhile, users can choose to draw the acquired PPG

waveform to visual the signal quality. After data acquisition is completed, each subject data will be stored and saved in the database, and the computer software supports the export of physiological information and PPG signal from this database to .csv file by subject ID number, which is convenient for offline data processing and analysis. The platform integrates the functions of data reception, real-time mapping, data storage and subject information management, which makes data easy to use and the system more practical. **Figure 1** illustrates the experimental design of this study, it mainly consists three parts: the PPG signal acquisition device, the working principle of PPG sensor, and acquired four-wavelength PPG waveform.

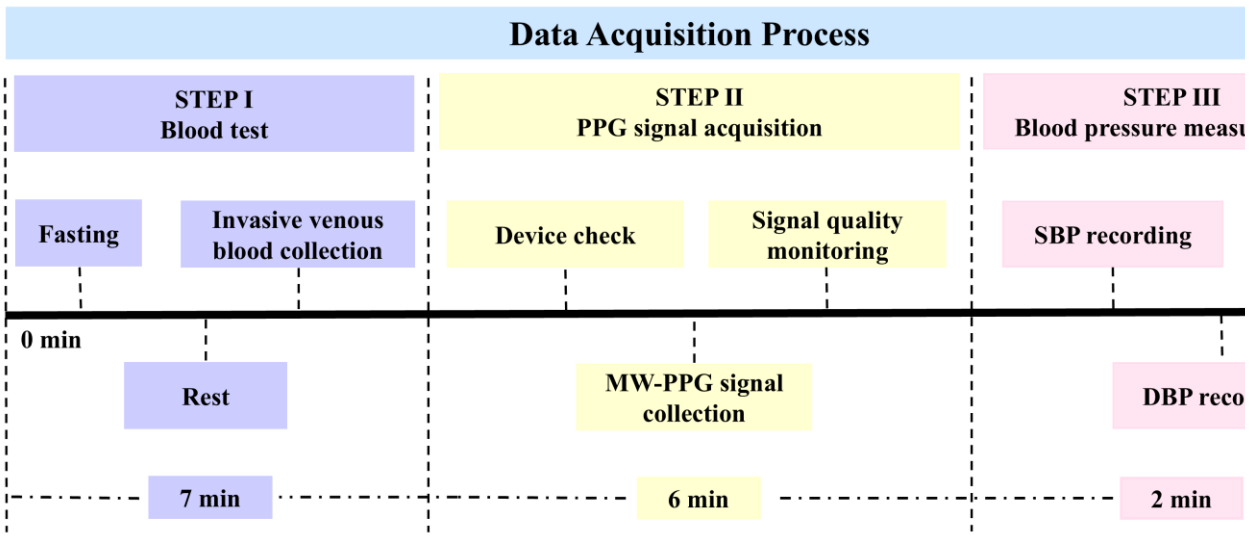

Figure 2. Illustration of data acquisition process. Data acquisition comprises invasive blood test, PPG signal acquisition and brachial artery blood pressure measurement three steps. The whole process takes about 15 minutes.

● Data acquisition

The entire data acquisition process takes about 15 minutes and consists invasive blood test, PPG signal acquisition and blood pressure measurement three steps, as shown in **Figure 2**. Hb concentration and blood glucose were acquired from venous blood samples. Subjects were required to cease eating before 9:00 p.m. on the day before the blood test for at least 8 hours, to obtain fasting blood glucose and ensures Hb and blood glucose are obtained under consistent clinical conditions[22]. During acquisition day, a venous blood sample was collected by medical staffs from the university hospital. As a reference, Hb concentration was measured by HemoCue Hb 201+ portable hemoglobin analyzer (HemoCue AB, Ängelholm, Sweden), it was calibrated against the hemiglobincyanide (HiCN) method, the international reference method recommended by the International Council for Standardization in Haematology (ICSH). The analyzer is factory calibrated and needs no further calibration[23]. Before usage, ensure it works with an appropriate environment temperature 15–30 degrees (°C), a humidity 40–80% and is fully charged. Press the left button until all symbols appear on the display. The analyzer performs a self-test to verify the performance of the

optronic unit[24]. Blood glucose was measured by auto hematology analyzer BC-6100 (Mindray, Co., Ltd., Shenzhen, China), researchers conduct calibration using the calibrator from manufacturer before experiment and ensure it works at 18–25 degrees (°C), 40–80% humidity[25]. In order to ensure data reliability, subjects were told to refrain from alcohol intake and intensive exercise at least 2 hours before the experiment[26]. Researchers inquiries about these criteria, only eligible participants were allowd to enter the data acquisition room. After entering data acquisition room, subjects were given 5 minutes to familiarize themselves with the environment and maintain relaxed and clam.

During PPG acquisition process, set the waveform sampling rate as 200 Hz, and the analog-to-digital conversion precision as 12-bit. Researchers collected PPG signals from the left index fingertip of subjects using the MW-PPG acquisition device developed in this study. The signal recording lasts for 60 seconds, and each subject comprises four PPG channels with each channel contains 12,000 sampling points. During PPG data acquisition, researchers monitored recorded signals through the computer software and ensured waveform was relative smooth and has discernible periods, otherwise, subjects needed to take a break and re-acquire. This process aims to minimize the interference of motion artifacts and ensure the quality of the acquired signals.

After PPG signal acquisition, we measured brachial arterial blood pressure of each subject using Omron U724J

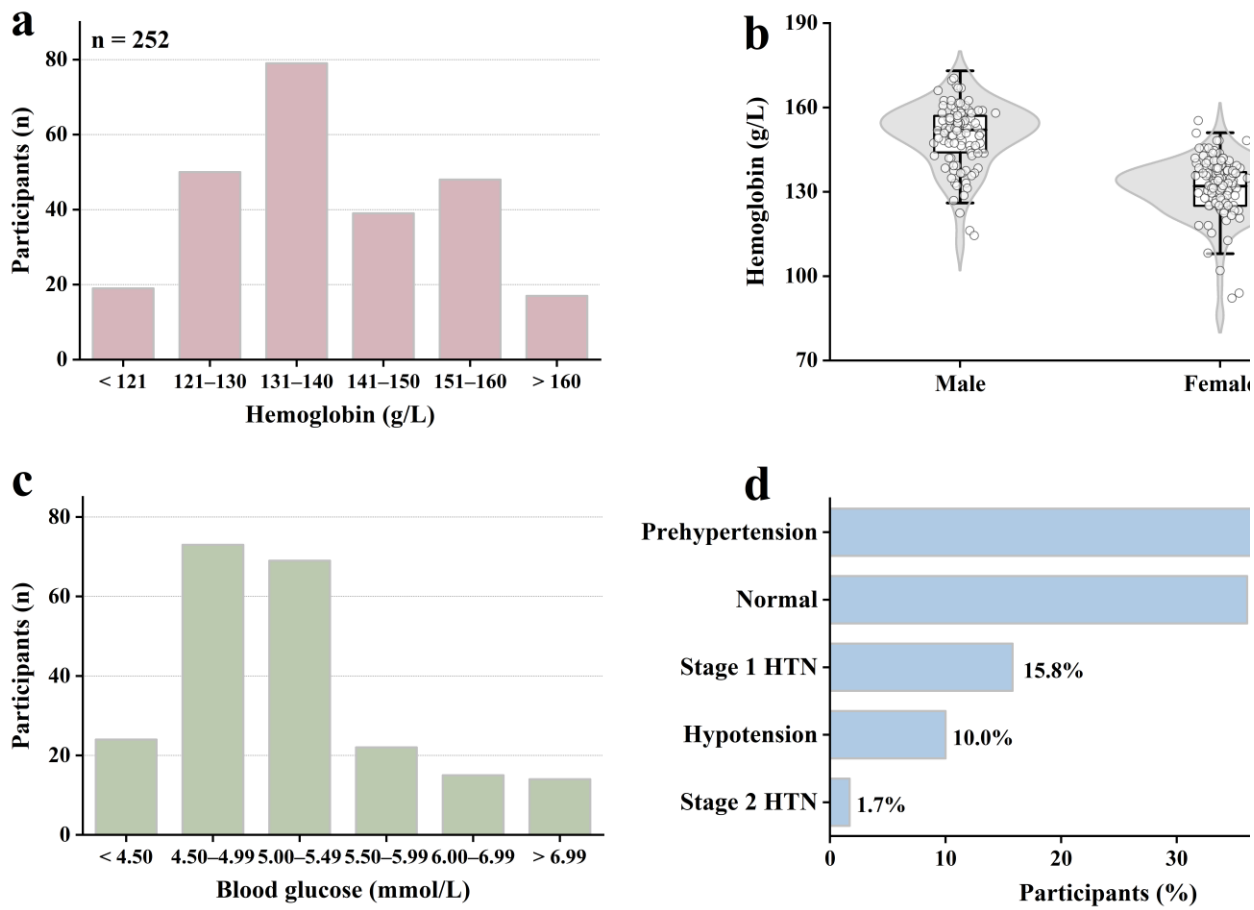

Figure 3. (a) Distribution of subject numbers in different hemoglobin concentration intervals. **(b)** Boxplots of hemoglobin levels for male and female subjects. **(c)** Distribution of subject numbers in different blood glucose concentration intervals. **(d)** Distribution of subjects in different blood pressure stages.

(Omron Corp., Kyoto, Japan) electronic upper arm sphygmomanometer[27]. Each participant was asked to sit upright with their legs uncrossed, and relax their right forearms on a clean table, ensuring they were naturally extended. We recorded systolic (SBP) and diastolic (DBP) blood pressure of subjects.

This study was approved by the medical Institutional Review Board (IRB) of Guilin University of Electronic Technology in China (approval number: GUET-20220301-001), and complied with regulations set by it. Each participant has been fully informed the experimental procedure and signed an informed consent before the study. All participants were compensated monetarily at 20 Yuan. Participants fill in the necessary information before data collection, including age, gender, height and weight, and data acquisition was conducted at a private, and comfortable clinical room in the university hospital.

● Participants

The dataset was acquired from 252 adult participants, participants were recruited through advertisements on university bulletin boards and personal recommendations. The age range of the subjects

is 21–90 years (mean±standard deviation age, 47±21 years), and the gender distribution is relatively balanced, with 56.7% females (109 male, 143 female). This dataset covered people with different Hb concentration levels, and recorded information on fasting blood glucose concentration and brachial artery blood pressure. Some subjects suffered from diabetes and cardiovascular diseases such as anemia and hypertension. The statistical results are shown in **Figure 3**.

● De-identification

In the process of creating the dataset, we removed personal information of each subject such as name, record date, personal address and so on, in order to protect their privacy. And named the data by ID plus the number, e.g. ID1, ID2, ..., ID252.

Data Records

The dataset has been packed and is available on Figshare research repository, it is under international (CC BY 4.0) license (https://doi.org/10.6084/m9.figshare.22256143.v6)[28]. The dataset includes a table file and 252 waveform data files. The table file is named "subject information.xlsx", it records basic information of all subjects, including age, gender, height, weight, Hb concentration, blood glucose concentration, systolic and diastolic blood pressure values. The ID number represents the number of subjects. "Signal length" column represents the four-wavelength PPG length in seconds (e.g., 60 s, 50 s, 45 s). PPG data was stored in .mat format. Each .mat file was named by subject ID, representing one subject PPG data, it contains a structure named "PPGdata" which includes four fields: nm_660, nm_730, nm_850, and nm_940. Each field is a column vector, these fields represent 660 nm, 730 nm, 850 nm and 940 nm wavelength PPG signal, respectively. For ease of access, four-wavelength PPG data is co-deposited as .csv format, named by subject ID, and the 660 nm, 730 nm, 850 nm and 940 nm wavelength PPG were stored in the first to fourth column, respectively.

Technical Validation

The MW-PPG signal quality has an impact on estimation accuracy of physiological parameters such as Hb concentration, and in order to ensure the reliability of PPG signal, validation of the waveform quality is particularly important. Therefore, we used signal-to-noise ratio (SNR) to assess the quality of acquired PPG signal in this section. It reflects the ratio of useful information of the signal to the degree of noise interference, and has been widely used while examining the signal validity[29,30].

The definition of SNR varies a little, but the primary object is to quantify the proportion between valid information and noise in one signal record. For example, Elgendi et al. used the ratio of signal variance to noise variance to define SNR and used it as a basis for validating the optimal signal quality index[31]. Parsaoran et al. utilized the ratio of the filtered signal to the signal noise to define SNR and used it as one of the key metrics for evaluating the denoising effects of PPG signals[32]. In this study, we used the ratio

between the bandpass signal and the bandstop signal in decibels (dB), defined as follows:

$$SNR=10\log_{10}\left(S_{Bandpass}/N_{Bandstop}\right)$$

Where $S_{Bandpass}$ indicates a bandpass signal in the range of 0.5 to 10 Hz and $N_{Bandstop}$ indicates a bandstop signal outside the range of 0.5 to 10 Hz.

To ensure the calculation accuracy of SNR, four channels of each subject signal were preprocessed to remove interfering factors such as baseline drift. Then, we used Chebyshev type II second-order filter to process each PPG channel since it can better remove high-frequency noise while maintaining the original shape of the signal, avoiding the loss of systolic peaks, dicrotic notches and other fiducial points[33]. After processing stage, the bandpass and bandstop signal are calculated and SNR is further figured. The validation algorithm operates on MathWorks MATLAB R2023b.

● Data screening

Before archiving the data acquired from the university hospital, it is necessary to go through data screening processes to eliminate abnormal data and correct defective data, so as to ensure high quality of the final dataset. The data screening process consists data integrity check, data availability check and signal quality assessment three steps, as shown in **Figure 4**. After data screening processes, we excluded 15 subjects data due to invalid signals, and finally retained valid data from 252 subjects. The final dataset contains 1008 PPG data segments with each subject contributing 4 valid wavelengths. The steps are described as follows:

1. Data integrity check: This process verifies for missing values for the following: subjects' basic physiological information (gender, age), hemoglobin values, and four-channel waveform data. If one or more items are missing for one subject, the corresponding record was discarded.

2. Data availability check: Examine noise in PPG, and ensure signal components such as systolic and diastolic phase are identifiable. This may caused by loose sensor connections, ambient light or industrial frequency interference. We removed distorted PPG segments with signal components unrecognizable, and if signal length for any channel was less than 30 seconds, the corresponding subject record was excluded.

3. Signal quality assessment: Check the signal quality using SNR. If SNR for all four PPG channels are less than 10 dB, the signal quality was considered poor, and the corresponding subject data was removed[34].

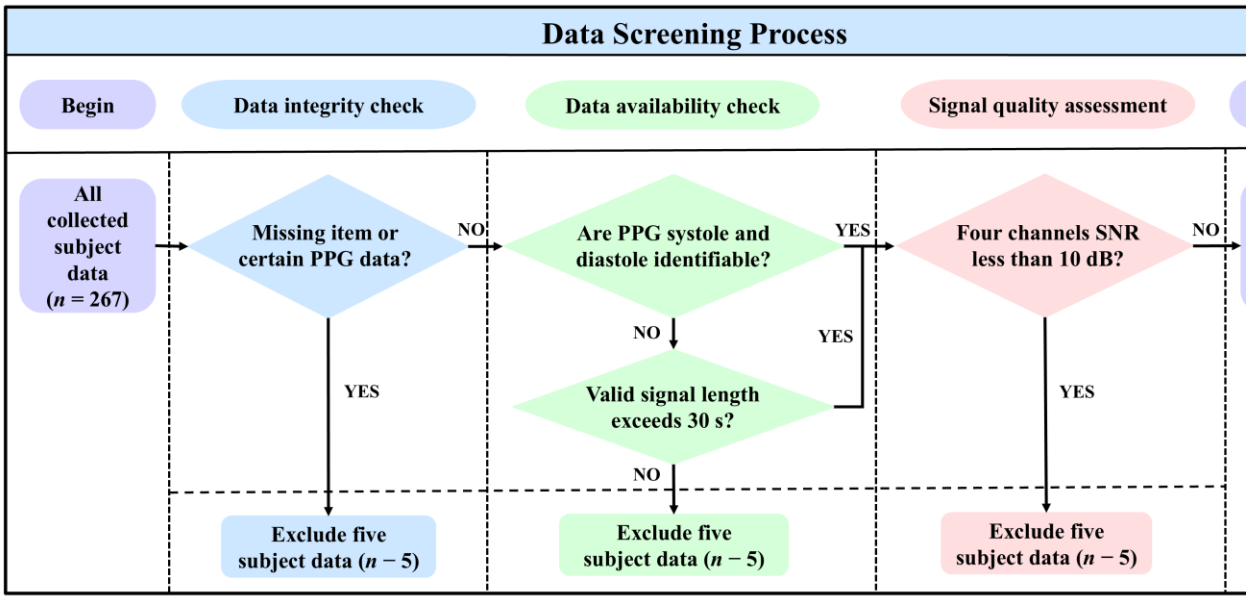

Figure 4. The flowchart of data screening process.

● Validation of four-wavelength PPG waveforms

The boxplots and bar charts of SNR statistical results for four wavelengths PPG are shown in **Figure 5**. It shows discrepancies of four PPG wavelengths, 660 nm and 850 nm signal exhibits better SNR values and stability, while 940 nm signal exhibits worse. For 850 nm, SNR has the highest mean SNR 19.04 dB, the lowest standard deviation 4.77 dB, and 41.7% subjects with SNR greater than 20 dB, while 0.8% subjects less than 10 dB. For 940 nm, SNR has the lowest mean 16.44 dB, the largest standard deviation 6.53 dB, and 34.5% subjects with SNR greater than 20 dB, 17.5% subjects less than 10 dB, indicating poor signal stability. Results for 730 nm are distributed between 850 nm and 940 nm.

Table 1, Table 1 summarizes the statistical power, effect size (ES), p-values, and confidence intervals (CI) used to assess statistical differences in SNR across wavelength conditions. Confidence intervals were estimated using the Bootstrap percentile method, and wavelength comparisons were performed using the Friedman test with Bonferroni correction applied to account for six pairwise comparisons[35]. All wavelength pairs exhibited statistically significant differences (corrected $p < 0.05$). Notably, the comparison between 730 nm and 940 nm achieved a statistical power of 0.990, indicating a 99% probability of detecting a true difference, while the largest effect size was observed between 850 nm and 940 nm (Cohen's $d = 0.742$). These results demonstrate systematic, wavelength-dependent differences in PPG signal quality, consistent with variations in optical penetration depth and wavelength-specific light attenuation within heterogeneous biological tissues[12].

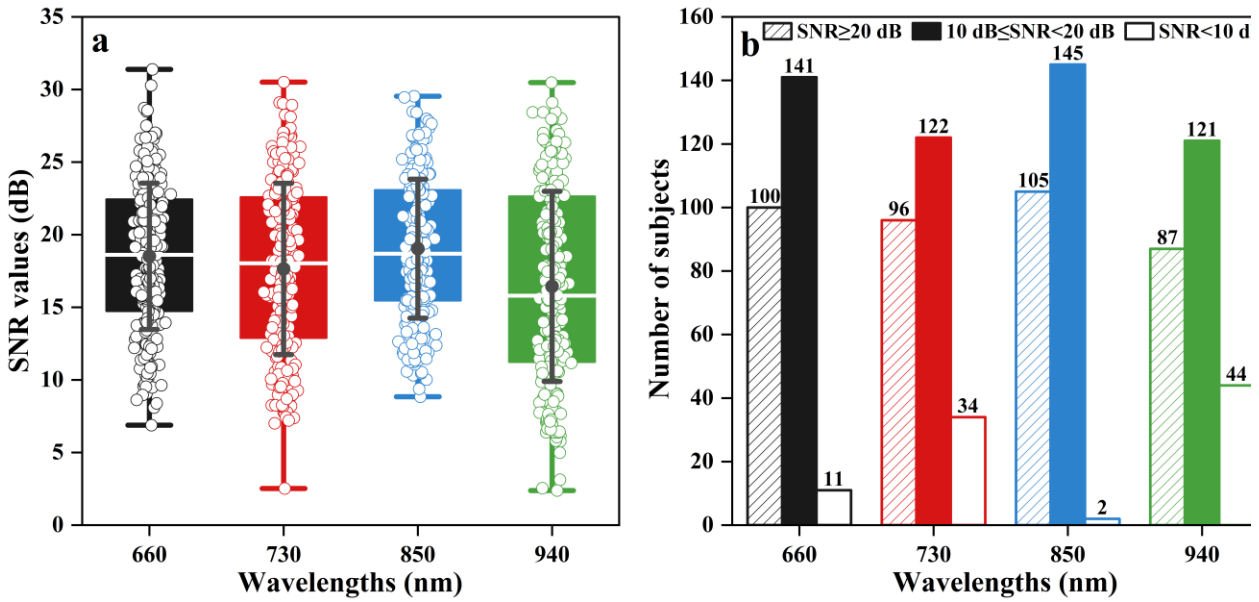

Figure 5. (a) Boxplot of SNR at 660 nm, 730 nm, 850 nm and 940 nm four wavelengths. **(b)** Distribution of subjects across three SNR categories, designated as SNR ≥ 20 dB, 10 dB ≤ SNR < 20 dB, and SNR < 10 dB, illustrated as 660 nm, 730 nm, 850 nm and 940 nm four wavelengths. The hatched pattern indicates SNR greater or equals to 20 dB, solid fill indicates SNR greater or equals to 10 dB and less than 20 dB, empty fill indicates SNR less than 10 dB.

Table 1. Statistical results of SNR for PPG data at 660 nm, 730 nm, 850 nm and 940 nm wavelengths. w1 and w2 (95% CI): using Bootstrap sampling for SNR confidence interval (CI) at one wavelength. Diff (w1 − w2), 95% CI: using Bootstrap sampling for differences between two wavelengths SNR means. Effect size (ES) was computed as average of paired SNR differences divided by standard deviation. *p* (raw): uncorrected *p*-value. *p* (corrected): *p*-value after Bonferroni correction. α: significance level (0.05). sample size=252.

Comparison (wavelengths)		w1 (95% CI)	w2 (95% CI)	Diff (w1 − w2), 95% CI	ES (Cohen's *d*)	Statistical power	*p* (raw)	*p* (corrected)
w1	w2							
660 nm	730 nm	(17.16, 19.24)	(16.67, 18.76)	(0.60, 1.15)	0.358	0.980	1.29×10^{-8}	7.73×10^{-8}
660 nm	850 nm	(17.16, 19.24)	(17.96, 19.64)	(−0.74, −0.31)	0.362	0.982	9.13×10^{-9}	5.48×10^{-8}
660 nm	940 nm	(17.16, 19.24)	(14.80, 17.32)	(1.72, 2.42)	0.635	1.000	6.78×10^{-24}	4.07×10^{-23}
730 nm	850 nm	(16.67, 18.76)	(17.96, 19.64)	(−1.76, −1.04)	0.451	0.999	7.68×10^{-13}	4.61×10^{-12}
730 nm	940 nm	(16.67, 18.76)	(14.80, 17.32)	(0.81, 1.59)	0.382	0.990	1.29×10^{-9}	7.74×10^{-9}
850 nm	940 nm	(17.96, 19.64)	(14.80, 17.32)	(2.29, 2.91)	0.742	1.000	4.71×10^{-32}	2.82×10^{-31}

In reflective PPG sensors, as the wavelength increases, the optical range length becomes longer and the average penetration depth increases, and different light wavelengths can acquire physiological information at different depths[36]. Human tissue exhibits varying absorption coefficients for different light

wavelengths[37], and the light will be affected by the interactions of absorption, reflection, and scattering (i.e., photon-tissue interactions) when it penetrates human

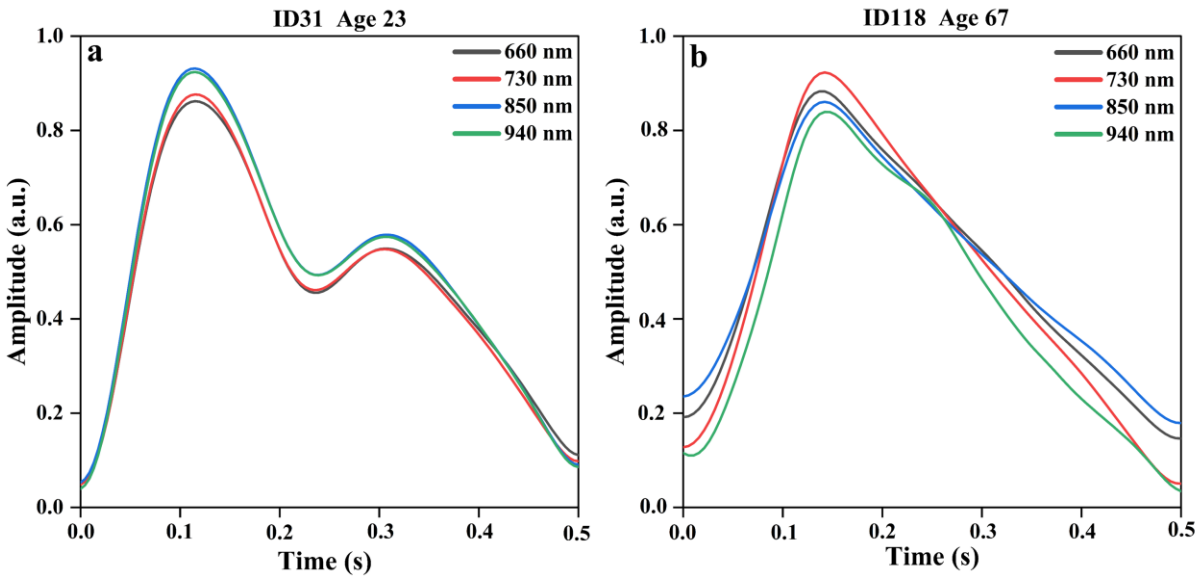

Figure 6. Example of one cycle PPG signal after filtering (Chebyshev II second-order bandpass filter, 0.5–10 Hz)[33] from subject ID31 and ID118 at 660 nm, 730 nm, 850 nm and 940 nm wavelengths, reflecting PPG morphological differences caused by depths of light penetration into tissue and subject age. **(a)** ID31 indicates subject number 31. **(b)** ID118 indicates subject number 118.

tissues. Thus, different light wavelengths exhibit distinctive optical properties, reflecting the corresponding physiological information[38]. The SNR characteristics of PPG signal are closely related to the optical behavior of Hb. The distinctive absorption properties of Hb at four distinct wavelengths (660 nm, 730 nm, 850 nm, 940 nm) indicate four-wavelength PPG signal is eligible to reflect oxygenation state and changes of Hb concentration (**Figure 6**). The percentages of subjects with SNR greater 10 dB at four wavelengths are over 80%. Through wavelength-dependent analysis, SNR provides a quantitative basis for signal quality assessment and reveals the potential used for Hb extraction and other physiological information. Statistical results illustrate discrepancies for different wavelengths. Therefore, SNR statistical results showed the proposed Hb-PPG dataset is reliable, and is eligible for discovering biochemical states in tissues and estimation of physiological parameters such as Hb, blood glucose and blood pressure.

Usage Notes

The Hb-PPG dataset was stored in standard file formats (mat, csv and xlsx), can be read by various software packages including MATLAB and R. facilitating data reading and processing. It covers a hemoglobin concentrations range of 85–173 g/L, along with diabetes, hypertension cases with cardiovascular diseases. Researchers can select different PPG wavelengths or choose data records under different SNR conditions for further analysis. If researchers use this repository or its components, please cite the publication as reference[28]. Specifically, (1) Multiple types of PPG features can be extracted,

including morphological features (e.g., time-, frequency-domain, etc.) and statistical features (e.g., mean, standard deviation, skewness, quartiles and other metrics) to analyze correlations among four-wavelength PPG signals, and disclose differences of MW-PPG in reflecting physiological information[39]; (2) Analyze the relation between MW-PPG characteristics (e.g., pulse transit time, PTT) and hemodynamics and blood fluid properties, and investigate the physiological mechanisms hemoglobin levels affects MW-PPG. Based on machine learning and deep learning techniques, establish classification, estimation models for hemoglobin, blood glucose or blood pressure[40,41,42] or perform joint research of multiple physiological parameters[43]; (3) It may helpful for conducting joint research or data collection on cardiac and brain imaging to investigate brain-heart associations, especially four-wavelength PPG and multi-frequency bands. Researches indicate multi-frequency imaging helps decoding complex brain activity, exhibiting similar to multi-wavelength optical signals and non-invasive Hb measurement. This may provide value for future collection of brain imaging dataset and studies in cardiovascular and neuroscience fields[44,45]; (4) Evaluate algorithm performances for Hb, blood glucose and blood pressure[46,47], and devise methods to help improve the accuracy and reliability of non-invasive hemoglobin testing and health monitoring devices[48,49].

Hb-PPG dataset minimize errors through data acquisition device and data screening process. Hemoglobin, blood glucose and blood pressure were acquired by HiCN testified and clinical hematology analyzer, and are unlikely change during experimental period. Thus, the collected data can be considered reliable doing test-retests[50,51]. However, skin tone and finger skin thickness may vary across different regions and body profiles, affecting blood perfusion and introducing physiological variations in PPG signals[52]. There also exists variances for hemoglobin in male and female, healthy adults and children, pregnants groups. Researchers should note this when applying it to different populations[53]. In addition, hemoglobin and blood glucose were obtained via venous blood samples using portable device and hematology analyzer, researchers should acknowledge this when comparing with other studies from fingerstick capillary sampling[54].

By analyzing the PPG waveform data, researchers were able to extract physiological indicators that are closely related to health status, contributing to non-invasive health monitoring, early detection of cardiovascular diseases and development of personalized treatment. Hb-PPG helps promote the application of PPG technique and enhance clinical and ambulatory health monitoring devices in public health.

References

1. Ahmed, M. H., Ghatge, M. S. & Safo, M. K. Hemoglobin: structure, function and allostery. *Subcell Biochem.* **94**, 345–382, https://doi.org/10.1007/978-3-030-41769-7_14 (2020).
2. Lee, G. *et al*. Association of hemoglobin concentration and its change with cardiovascular and all-

cause mortality. *J Am Heart Assoc.* **7**, e007723, https://doi.org/10.1161/JAHA.117.007723 (2018).

3. Benson, C. S. *et al*. The effect of iron deficiency and anaemia on women's health. *Anaesthesia* **76**, 84–95, https://doi.org/10.1111/anae.15405 (2021).
4. Mantadakis, E., Chatzimichael, E. & Zikidou, P. Iron deficiency anemia in children residing in high and low-income countries: risk factors, prevention, diagnosis and therapy. *Mediterr J Hematol Infect Dis.* **12**, e2020041, https://doi.org/10.4084/mjhid.2020.041 (2020).
5. Qin, T. T. *et al*. Association between anemia and cognitive decline among Chinese middle-aged and elderly: evidence from the China health and retirement longitudinal study. *BMC Geriatr.* **19**, 1–13, https://doi.org/10.1186/s12877-019-1308-7 (2019).
6. Karakochuk, C. D. *et al*. Measurement and interpretation of hemoglobin concentration in clinical and field settings: a narrative review. *Ann N Y Acad Sci.* **1450**, 126–146, https://doi.org/10.1111/nyas.14003 (2019).
7. Li, G. *et al*. Non-invasive detection of haemoglobin, platelets, and total bilirubin using hyperspectral cameras. *Talanta* **260**, 124634, https://doi.org/10.1016/j.talanta.2023.124634 (2023).
8. Dimauro, G., Caivano, D. & Girardi, F. A new method and a non-invasive device to estimate anemia based on digital images of the conjunctiva. *IEEE Access* **6**, 46968–46975, https://doi.org/10.1109/ACCESS.2018.2867110 (2018).
9. Lychagov, V. V. *et al*. Noninvasive hemoglobin measurements with photoplethysmography in wrist. *IEEE Access* **11**, 79636–79647, https://doi.org/10.1109/ACCESS.2023.3300293 (2023).
10. Acharya, S., Swaminathan, D., Das, S., Kansara, K. & Aatre, K. R. Non-Invasive Estimation of Hemoglobin Using a Multi-Model Stacking Regressor. *IEEE J Biomed Health Inform.* **24**, 1717–1726, https://doi.org/10.1109/JBHI.2019.2954553 (2019).
11. Kavsaoğlu, A. R., Polat, K. & Hariharan, M. Non-invasive prediction of hemoglobin level using machine learning techniques with the PPG signal's characteristics features. *Appl. Soft Comput* **37**, 983–991, https://doi.org/10.1016/j.asoc.2015.04.008 (2015).
12. Ray, D., Collins, T., Woolley, S. I. & Ponnapalli, P. V. A review of wearable multi-wavelength photoplethysmography. *IEEE Rev Biomed Eng.* **16**, 136–151, https://doi.org/10.1109/RBME.2021.3121476 (2023).
13. Park, J., Seok, H. S., Kim, S. S. & Shin, H. Photoplethysmogram analysis and applications: an integrative review. *Front Physiol.* **12**, 808451, https://doi.org/10.3389/fphys.2021.808451 (2022).
14. Tamura, T. Current progress of photoplethysmography and SPO2 for health monitoring. *Biomed Eng Lett.* **9**, 21–36, https://doi.org/10.1007/s13534-019-00097-w (2019).
15. Jia, Y. L., Zheng, Y., Wang, Z. & Li, Y. Non-invasive blood glucose based on dual wavelength transmission PPG signal. *IEEE Sens. J.* **25**, 11876–11884, https://doi.org/10.1109/JSEN.2025.3539604 (2025).
16. Salah, M. *et al*. Beat-based PPG-ABP cleaning technique for blood pressure estimation. *IEEE Access* **10**, 55616–55626, https://doi.org/10.1109/ACCESS.2022.3175436 (2022).
17. Ouyang, V. *et al*. The use of multi-site photoplethysmography (PPG) as a screening tool for coronary arterial disease and atherosclerosis. *Physiol Meas.* **42**, 064006, https://doi.org/10.1088/1361-6579/abad48 (2021).
18. Friebel, M., Helfmann, J., Netz, U. & Meinke, M. Influence of oxygen saturation on the optical scattering properties of human red blood cells in the spectral range 250 to 2000 nm. *J Biomed Opt.* **14**, 034001, https://doi.org/10.1117/1.3127200 (2009).

19. Faber, D. J. *et al*. Oxygen saturation-dependent absorption and scattering of blood. *Phys Rev Lett.* **93**, 028102, https://doi.org/10.1103/PhysRevLett.93.028102 (2004).
20. Spigulis, J., Gailite, L., Lihachev, A. & Erts, R. Simultaneous recording of skin blood pulsations at different vascular depths by multiwavelength photoplethysmography. *Appl Opt.* **46**, 1754–1759, https://doi.org/10.1364/AO.46.001754 (2007).
21. Wang, Y. Y, Li, G., Kong, L. & Lin, L. High-precision non-invasive RBC and HGB detection system based on spectral analysis. *Anal Bioanal Chem.* **415**, 6733–6742, https://doi.org/10.1007/s00216-023-04950-x (2023).
22. Whitehead, R. D., Mei, Z. G., Mapango, C. & Jefferds, M. E. D. Methods and analyzers for hemoglobin measurement in clinical laboratories and field settings. *Ann N Y Acad Sci.* **1450**, 147–171, https://doi.org/10.1111/nyas.14124 (2019).
23. HemoCue Hb 201+ System. https://hemocue.com/en/products/hematology/hemocue-hb-201plus-system/ (Online).
24. Namaste, S. M., Baingana, R. & Brindle, E. Hemoglobin measurement in venous blood compared with pooled and single-drop capillary blood: a method-comparison study in a controlled and survey setting in Uganda among children and women. *Am J Clin Nutr.* **119**, 949–959, https://doi.org/10.1016/j.ajcnut.2023.12.025 (2024).
25. Mindray Auto Hematology Analyzer BC-6000. https://www.mindray.com/en/products/hematology (Online).
26. Kujovich, J. L. Evaluation of Anemia. *Obstet Gynecol Clin North Am.* **43**, 247–64, https://doi.org/10.1016/j.ogc.2016.01.009 (2016).
27. Lu, Y. F. *et al.* Association between plasma and dietary trace elements and obesity in a rural Chinese population. *Br J Nutr.* **131**, 123–133, https://doi.org/10.1017/S0007114523001435 (2023).
28. Chen, L. Q., Li, S. Y., Liang, Y. B., Chen, Z. C. & Elgendi, M. *Figshare*, https://doi.org/10.6084/m9.figshare.22256143.v6 (2025).
29. Blanco-Velasco, M., Weng, B. & Barner, K. E. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Comput. Biol. Med* **38**, 1–13, https://doi.org/10.1016/j.compbiomed.2007.06.003 (2008).
30. Smital, L. *et al*. Real-time quality assessment of long-term ECG signals recorded by wearables in free-living conditions. *IEEE Trans Biomed Eng.* **67**, 2721–2734, https://doi.org/10.1109/TBME.2020.2969719 (2020).
31. Elgendi, M. Optimal Signal Quality Index for Photoplethysmogram Signals. *Bioengineering* **3**, 21, https://doi.org/10.3390/bioengineering3040021 (2016).
32. Parsaoran, A. J., Mandala, S. & Pramudyo, M. Study of denoising algorithms on photoplethysmograph (PPG) signals. *In 2022 IEEE International Conference on Data Science and Its Applications (ICoDSA)*, 289–293, https://doi.org/10.1109/ICoDSA55874.2022.9862918 (2022).
33. Liang, Y. B., Elgendi, M., Chen, Z. C. & Ward, R. An optimal filter for short photoplethysmogram signals. *Sci. Data* **5**, 1–12, https://doi.org/10.1038/sdata.2018.76 (2018).
34. Garcia-Casal, M. N., Pasricha, S. R., Sharma, A. J. & Pena-Rosas, J. P. Use and interpretation of hemoglobin concentrations for assessing anemia status in individuals and populations: results from a WHO technical meeting. *Ann N Y Acad Sci.* **1450**, 5–14, https://doi.org/10.1111/nyas.14090 (2019).
35. Serdar, C. C., Cihan, M. Yucel, D. & Serdar, M. A. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochem Med*

(Zagreb). **31**, 010502, https://doi.org/10.11613/BM.2021.010502 (2021).

36. Savo, R., Pierrat, R., Najar, U., Carminati, R., Rotter, S. & Gigan, S. Mean path length invariance in multiple light scattering. *Science* **358**, 765–768, https://doi.org/10.1126/science.aan4054 (2017).
37. Chatterjee, S., Budidha, K. & Kyriacou, P. A. Investigating the origin of photoplethysmography using a multiwavelength Monte Carlo model. *Physiol Meas.* **41**, 084001, https://doi.org/10.1088/1361-6579/aba008 (2020).
38. Jacques, S. L. Optical properties of biological tissues: a review. *Phys Med Biol.* **58**, R37, https://doi.org/10.1088/0031-9155/58/11/R37 (2013).
39. Sidorov, I. S., Romashko, R. V., Koval, V. T., Giniatullin, R. & Kamshilin, A. A. Origin of infrared light modulation in reflectance-mode photoplethysmography. *PLoS One* **11**, e0165413, https://doi.org/10.1371/journal.pone.0165413 (2016).
40. Chakraborty, S., Kansara, K., Kumar, R. D., Swaminathan, D., Aatre, K. & Acharya, S. Non-invasive estimation of clinical severity of anemia using hierarchical ensemble classifiers. *J Med Biol Eng.* **42**, 828–838, https://doi.org/10.1007/s40846-022-00750-3 (2022).
41. Habbu, S., Dale, M. & Ghongade, R. Estimation of blood glucose by non-invasive method using photoplethysmography. *Sādhanā* **44**, 135, https://doi.org/10.1007/s12046-019-1118-9 (2019).
42. Liu, J. *et al*. PCA-based multi-wavelength photoplethysmography algorithm for cuffless blood pressure measurement on elderly subjects. *IEEE J Biomed Health Inform.* **25**, 663–673, https://doi.org/10.1109/JBHI.2020.3004032 (2020).
43. Monte-Moreno, E. Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. *Artif Intell Med.* **53**, 127–138, https://doi.org/10.1016/j.artmed.2011.05.001 (2011).
44. Gong, Z., Q. & Zuo, X. N. Dark brain energy: Toward an integrative model of spontaneous slow oscillations. *Phys Life Rev.* **52**, 278–297, https://doi.org/10.1016/j.plrev.2025.02.001 (2025).
45. Song, J., Li, C. Q. & Chansanam, W. Construction of medical scientific data repositories in China: analysis of survey and recommendations. *Front Artif Intell.* **8**, 1544200, https://doi.org/10.3389/frai.2025.1544200 (2025).
46. Long, N. M. H. & Chung, W. Y. Wearable Wrist Photoplethysmography for optimal monitoring of vital signs: A unified perspective on pulse waveforms. *IEEE Photonics J.* **14**, 1–17, https://doi.org/10.1109/JPHOT.2022.3153506 (2022).
47. Karolcik, S., Ming, D. K., Yacoub, S., Holmes, A. H. & Georgiou, P. A Multi-Site, Multi-Wavelength PPG Platform for Continuous Non-Invasive Health Monitoring in Hospital Settings. *IEEE Trans Biomed Circuits Syst.* **17**, 349–361, https://doi.org/10.1109/TBCAS.2023.3254453 (2023).
48. Sabith, N. U. S., Sridevi, P., Arefin, K. Z., Rabbani, M. & Ahamed, S. I. Towards a Multi-model Comparative Study for Non-invasive Hemoglobin Level Prediction from Fingertip Video. *2024 IEEE International Conference on Digital Health (ICDH)*. 172–180, https://doi.org/10.1109/ICDH62654.2024.00038 (2024).
49. Song, Y. L. Non-invasive hemoglobin level estimation using photoplethysmography with performance comparison of advanced ML algorithms. *Signal Image Video P.* **19**, 446, https://doi.org/10.1007/s11760-025-04040-8 (2025).
50. Xu, T., Kiar, G., Zuo, X. N., Vogelstein, J. T. & Milham, M. P. Challenges in measuring individual differences of brain function. Imaging Neurosci (Camb). **3**, *imag_a_00430*, https://doi.org/10.1162/imag_a_00430 (2025).

51. Zuo, X. N., Xu, T. & Milham, M. P. Harnessing reliability for neuroscience research. *Nat Hum Behav.* **3**, 768–771, https://doi.org/10.1038/s41562-019-0655-x (2019).
52. Karakochuk, C. D. et al. Measurement and interpretation of hemoglobin concentration in clinical and field settings: a narrative review. *Ann N Y Acad Sci.* **1450,** 126–146, https://doi.org/10.1111/nyas.14003 (2019).
53. Anaemia. https://www.who.int/data/nutrition/nlis/info/anaemia (Online).
54. Killilea, D. W., Kuypers, F. A., Larkin, S. K. & Schultz, K. Blood draw site and analytic device influence hemoglobin measurements. *PLoS One.* **11**, e0278350, https://doi.org/10.1371/journal.pone.0278350 (2022).

Data Availability

The Hb-PPG dataset generated and analyzed during the current study is publicly available in the Figshare repository under a Creative Commons Attribution 4.0 International (CC BY 4.0) license at https://doi.org/10.6084/m9.figshare.22256143.v6. The repository includes all multi-wavelength photoplethysmography signal recordings and associated de-identified reference measurements required to reproduce the results reported in this work.

Code availability

The code used for technical validation in this study is accessible via GitHub repository (https://github.com/CardioWorks/Hemoglobin-Dataset-2025-main).

Acknowledgements

This study was supported in part by the National Natural Science Foundation of China (grant number 62361013, 62101148 and U22A2092), and Science and Technology Project of Guangxi Guike AD23023002. M. E. was funded by Khalifa University (grant number FSU-2025-001).

Author contributions

L. C., S. L., and L. L. designed the data acquisition equipment and system, conducted the data acquisition process, and wrote the manuscript. Y. Z. provided necessary hardware. G. H. helped acquire the clinical data. Prof. Z. C. and Y. L. supervised the project, with Y. L. assisted data acquisition. Prof. M. E. led the data processing and technical validation, providing valuable suggestions for improvement of the manuscript. All authors had read and approved the final manuscript.

Competing interests

The authors declare no competing interests. M. E. serves as Associate Editor for npj biosensing and had no role in the peer-review or decision to publish this manuscript.