



Towards non-invasive blood glucose measurement using machine learning: An all-purpose PPG system design



Shantanu Sen Gupta, Tae-Ho Kwon, Shifat Hossain, Ki-Doo Kim *

School of Electrical Engineering, Kookmin University, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
 Photoplethysmography
 Biomedical device
 Glucose
 Diabetes
 Features
 Machine learning

ABSTRACT

Diabetes, the result of excessive or uncontrolled glucose in the blood, is one of the leading causes of human mortality. Due to the unavailability of non-invasive glucose level checker until now, the most trustworthy day-to-day life glucose test for personal healthcare is the use of glucometer in which case painful finger pricking is an obvious part. However, researches have been done to prove the usage of pulse oximeter to measure the blood glucose level besides other physiological indicators such as heart rate, percentage of blood oxygen, etc. Here, as the first of two studies, we try to develop an all-purpose commercial prototype photoplethysmography (PPG) system to monitor necessary health indicator parameters in a non-invasive way. The developed fingertip PPG device consists of both transmissive and reflective type data acquisition system after illuminating the skin with red, green, and IR LEDs. Next, as the second study, special consideration is given to prove the efficiency of the device for measuring blood glucose level (BGL). To measure blood glucose from PPG signal, a few discriminative and related features are extracted from the obtained PPG signals. Machine learning algorithms are employed to predict the actual value of BGL from the extracted features. The proposed algorithm and system can predict the BGL level with a level of clinical accuracy. In the Clarke error grid plot, 96.15% and 3.85% of data are in the zone A and zone B, respectively, with 0% data in the critical zones.

1. Introduction

In modern days, measurement of human physiological parameters in a non-invasive way is becoming popular. For this purpose, the PPG system is gaining popularity among the medical research community and also individuals for their health care. The main reasons for adopting the PPG system is that it is wearable, easy to build, and requires a few hardware components. Previously, the PPG signal was reliably used to measure heart rate (HR), percentage of oxygen in blood (SpO_2), and breathing rate (BR), but other clinical measurements like blood glucose, blood pressure etc. are still in research phase (rarely commercially available). Moreover, PPG devices are subject to motion artifacts. So, obtaining a clean PPG signal by removing artifacts from the signal is also a challenge.

The word “PPG” was originated from the Greek “plethysmos” which means increasing and “graphein” which means to write. Therefore, combining these two words we obtain “Photoplethysmogram”. PPG signal records the change in blood volume over time. For this purpose, a light-emitting diode (LED) and photodetector (PD) setup is needed. The

LED illuminates the skin and the PD receives the transmitted or reflected signal depending on its location. The main layers of human skin are epidermis, dermis, subcutaneous tissue, and muscle. In each cardiac cycle when the heart pumps blood, a pressure pulse is felt by the arteries even though the pulse is somewhat damped when reaches the skin. This artery pulse is known as the primary peak of a cycle. Another peak, called the secondary peak is observed due to pressure pulse in the vein. These pulses can be detected by skin illumination and observing the changes in blood amount by PD. Actually, the PPG signal can be decomposed into two parts: AC and DC. The DC part is due to the non-pulsatile components under the skin, such as subcutaneous tissue, fat, bones, etc. Other contributing factors to DC part are respiration, thermoregulation, etc. However, overall, the origin of the PPG signal is not quite clear yet [1].

The pulse oximeter device which collects PPG signal can be either reflective or transmissive type. In case of “reflective” type, the PD and LED are on the same side, and for “transmissive” type, the PD and LED are on opposite side. Popular measurement sites in the human body are fingertip and wrist. Sometimes, the earlobe is also used as potential

* Corresponding author at: Dept. of Electronics Engineering, Kookmin University, 861-1 Jeongneung-dong, Seongbuk-gu, Seoul, 136-702, Republic of Korea.
 E-mail address: kdk@kookmin.ac.kr (K.-D. Kim).

location.

A relatively small amount of glucose found in the human blood is known as blood sugar level or blood glucose level. Although our body tightly regulates the level of blood glucose, sometimes malfunction of several organs and our food intakes cause hyperglycemia (high BGL) or hypoglycemia (low BGL). In the past decades, diabetes mellitus has been considered one of the main mortality causes. Among the two types of diabetes, type-1 is due to the failure of human body to produce insulin, and inability to use the produced insulin causes type-2 diabetes. As no remedy has yet been found to cure diabetes satisfactorily, keeping it under control is the best solution till now. To do that, knowing the level of glucose is most necessary. Among the most widely used and clinically accepted methods, besides clinical testing, invasive type glucose measurement by drawing blood is available. However, there is always a need to prick the finger which may lead to infection, and also the clinical test is not always available. So, to monitor the blood glucose continuously, a non-invasive PPG device development is urgent. Taking these facts into account, our current study aims to measure blood glucose levels in a non-invasive way through a device developed by ourselves.

In this study, our first target is to develop a reliable and compact PPG system in which induced noise is minimized and therefore recovering the original signal with minimal post-processing can be sufficient. Another main goal is to develop a blood glucose level (BGL) measurement system using the developed PPG device. To measure BGL, a set of PPG signal features that are closely related to blood glucose is first extracted and then machine learning (ML) models are used on top of it. The whole system is evaluated and justified against real-world cases.

The next sections are organized as follows: Section II will discuss the prior state-of-the-art researches in the related field. Our developed PPG system along with its uses in BGL estimation will be described in section III. Section IV will analyze the performance and limitations of this study in detail and, at last section V concludes the article.

2. Related works

From the last decade, personal healthcare devices and management have become popular. Two important aspects of this kind of device are wearable and clinically accurate measurements. In [2], Castaneda et al. discussed several cases of how PPG devices can be used as a wristwatch type, earring type, or forehead type device. Athavale et al. [3] also discussed the opportunities and challenges in the field of wearable health technology. Specifically, their focus was on electromyography (EMG), electroencephalogram (EEG), and electrocardiogram (ECG) devices and their performance analysis.

Poh et al. [4] developed a wearable PPG device for the earlobe to detect heart rate. Their IR LED-based reflective system consists of an accelerometer to further remove the motion artifacts and compared them to established ECG devices. Lee et al. [5] proposed another wearable PPG array sensor with IR LEDs as wristwatch. They showed that their developed device could achieve about 91% similarity to a commercial wristwatch PPG device for the data obtained from ulnar artery and radial artery. Another reflective type wristwatch was developed by Lee et al. [6] where two signals were collected from green LEDs. They also developed motion artifact removal techniques for four different cases (white noise, color noise, random hand movement vertically and horizontally). For SpO₂ measurement, Rodrigues et al. [7] developed a PPG device based on commercially available system on chip (SoC). Their established device consists of red and IR LED and collects data in a reflective mode. However, that device is not miniature and also not wearable. Recently, Banik et al. [8] built a flexible circuit board for reflective type PPG system. The developed wearable PCB successfully achieved the purpose of calculating heart rate and SpO₂.

Moreno et al. [9] measured BGL and blood pressure simultaneously from the PPG signal by means of machine learning algorithm. For both measurements, they used same type of statistical and signal energy-based features. Before getting into the feature extraction module,

each signal frame had to be passed by the activity detection module according to their method. In [10], Moreno et al. proposed another machine learning-based diabetes screening method. In that research, their main goal was to classify diabetes and non-diabetes. For this purpose, a set of different features such as low-frequency to high-frequency ratio of energies, zero-crossing rate, surrogate features, etc. were extracted. However, many demographic features such as gender, age, etc. have been used in their study. A near-infrared (NIR) PPG acquisition device prototype was designed by Hina et al. [11]. Moreover, the authors also tested their device for blood glucose estimation. But, their developed device was designed solely for estimating blood glucose levels. Another approach based only on the mean value of the acquired PPG signal is proposed by Zhou et al. [12]. In their approach, the PPG signal was considered as a summation of transient response, additive white Gaussian noise (AWGN), and clean signal. The first two components (transient response and AWGN) were removed in the first step and then the mean of clean signals was estimated. Finally, the BGL was estimated by L₁ norm.

In order to measure blood glucose level from smartphone spectroscopy signal, Dantu et al. [13] proposed an extended Beer-Lambert law for two different wavelengths (blue and green). Finally, they related the intensity ratio of the lights with actual blood glucose levels. Although they claimed that their system is fully dependent on the smartphone system, their setup needs an extra laser beam as that is only a transmissive system. Moreover, blue and green lights are far from the region where glucose shows the maximum absorbance. Another transmissive PPG system was developed by Paul et al. [14]. The authors used several voltage-related features of the signal to find the relation with blood glucose level. In that experiment, IR LED was used. Targeting to estimate the BGL, two Monte Carlo (MC)-based models were developed including the interaction of light and skin properties (bio-optical) by Liu et al. [15]. Their proposed models were Back Propagation Monte Carlo-Difference Equation Estimator (BpMC-DEE) and BpMC convolutional neural network (BpMC-CNN). Four different wavelengths (1850 nm, 1710 nm, 1520 nm, 1200 nm) were propagated through the earlobe and therefore five bio-optical properties (absorption coefficient, scattering coefficient, anisotropic factor, refractive index, tissue layer thickness) were obtained. On-line motion artifact removal and sensing blood glucose level estimation on field programmable gate array (FPGA) board was done at [16] by Ramasahayam et al. They obtained PPG signals from three different wavelengths (935 nm, 950 nm, 1070 nm) and calculated optical density from the signals. Later, optical density was used as a feature of artificial neural network (ANN). But the acquired PPG signal was very noisy. Consequently, the proposed noise cancellation algorithm did not perform well (the diastolic peaks were not clear). A CNN-based BGL estimation procedure was proposed most recently at [17]. The authors used pure raw signal after optimal filtering to feed to the deep learning architecture. Islam et al. [18] used a deep learning model to estimate BGL from raw PPG signal. To check other functional deviations due to glucose level, along with PPG data they also gave galvanic skin response (GSR) data to the deep learning model. Another deep learning-based diabetes classification work was done by Avram et al. [19]. In their research, it was shown that the deep neural network (DNN) score, after applying on the smartphone PPG waveform was strong enough to predict the diabetes independent of other factors such as age, body mass index (BMI), race, gender, etc. However, DNN score with comorbidities showed the highest area under receiver operating characteristic (ROC) curve (0.83).

Zhang et al. [20] developed a computational model for calculating BGL non-invasively based on energy conservation method. They combined several human physiological data such as HR, SpO₂, blood flow velocity, and basal metabolic rate. At the same time, they also developed a hardware to collect data. Their hardware system consisted of dual-wavelength (red and IR LEDs) transmissive type PPG acquisition, temperature sensor, and a radiation sensor. Habbu et al. [21] proposed a neural network model to detect the BGL. In their method, they used two

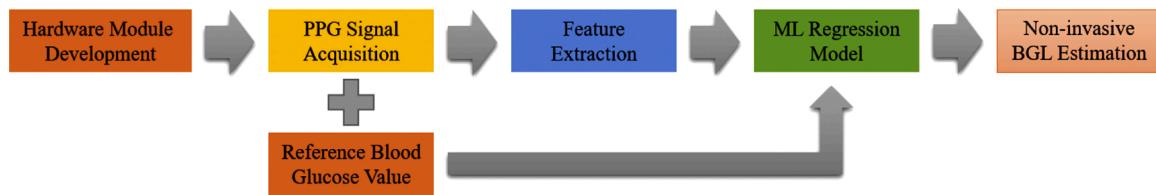


Fig. 1. Proposed system architecture at a glance.

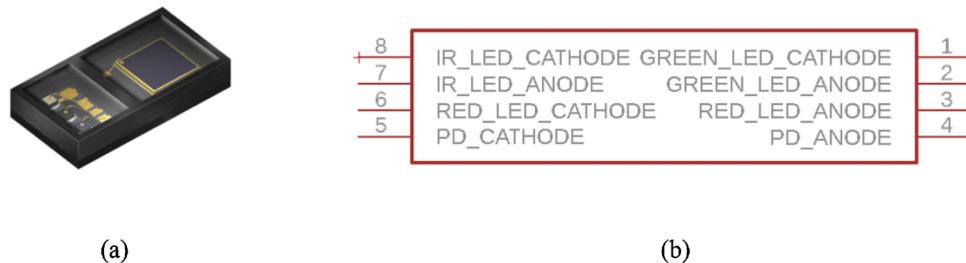


Fig. 2. SFH 7050 module. (a) physical SMD type, and (b) pinout diagram.

different feature sets: single pulse analysis (SPA), and frequency and time-based features. Their acquired PPG signal has one wavelength. Another single wavelength-based BGL estimation system was developed by Hina et al. [22]. In their work, they used NIR LED-based transmissive PPG signal and realized their model on FPGA. The PPG signal from video, recorded with a smartphone device was investigated by Chowdhury et al. [23]. The method [24] of Li et al. can be used to detect the level of severity of peripheral vascular occlusion due to type 2 diabetes. In their research, support vector machine was used for classification along with wolf pack search (WPS) algorithm.

The main focuses and our contributions in this research are as follows:

- A commercial level compact PPG device is developed.
- Proposed PPG device supports both reflective and transmissive type signal acquisition.
- The fingertip PPG device utilizes three different wavelengths: green, red, and IR.

- A machine learning based glucose estimation process is developed and evaluated by using the developed device.

3. Proposed method

In our proposed system, we develop an end to end scheme for blood glucose estimation. It includes the hardware system development to the final blood glucose level estimation. Referring to Fig. 1, the whole scheme can be considered as a summary of three separate modules: 1) hardware module, 2) data acquisition module, and 3) software processing module. A detailed description is given in this section for every segment of the system.

3.1. Hardware module

The developed hardware part here consists of two independent PPG signal attainment segments: reflective type PPG signal, and transmissive type PPG signal. However, the processor unit is unique in both cases.

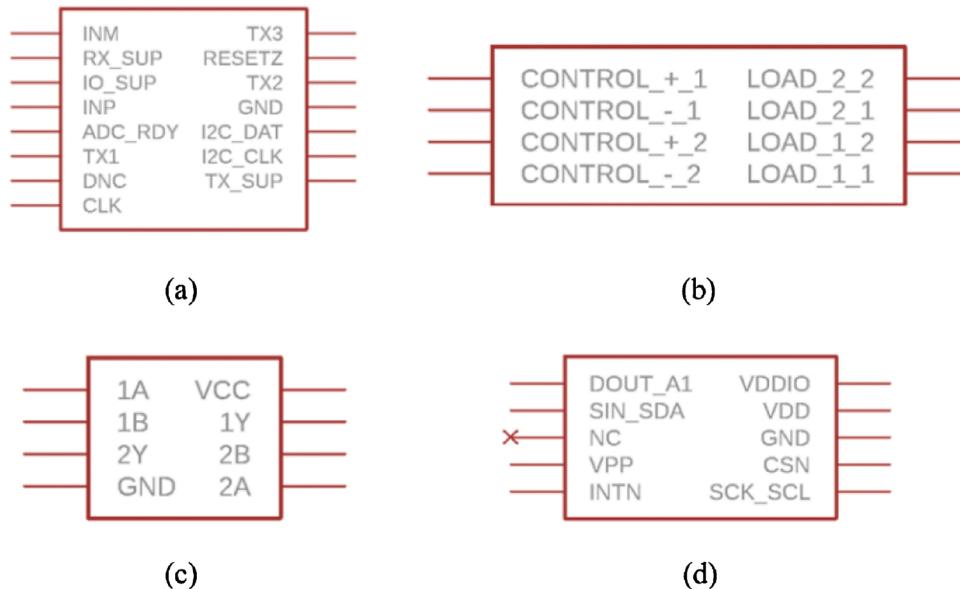


Fig. 3. Pinout diagram of (a) AFE4404, (b) CPC2030N, (c) 74AUP2G00DC-Q100H, and (d) MC365.

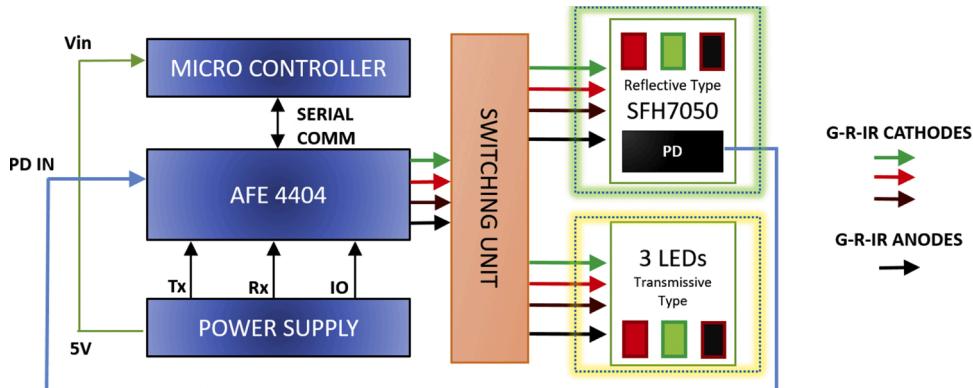


Fig. 4. Proposed hardware system as a block diagram.

The ESP32-PICO-V3 [25] is employed to control the entire system. This microcontroller has an on-chip RF communication system such as Wi-Fi and Bluetooth. Therefore, data can be easily transmitted to a remote server without any external communication module. The other significant features of this chip are 520 KB static random access memory (SRAM), dual-core 32-bit LX6 processor, maximum frequency 240 MHZ, touch sensor, I²S, I²C protocol, and general purpose input output (GPIO) pins. This SoC has a requirement of operating voltage between 3.0~3.6 V.

3.1.1. Reflective type PPG

For reflective type PPG system, SFH 7050 [26] surface mounted device (SMD) module is used. This module contains three different light emitting diodes (LEDs) and one photodetector (PD). The specifications of LEDs are: infrared ($\lambda_{peak} \rightarrow 950$ nm), red ($\lambda_{peak} \rightarrow 660$ nm), and green ($\lambda_{peak} \rightarrow 525$ nm). The PD can respond to the 400~1100 nm spectral range of sensitivity and has a very low dark current (~1 nA). Fig. 2(a) shows the SMD appearance of the SFH 7050 module, and Fig. 2(b) shows the schematic pinout diagram of the reflective type PPG sensing module.

3.1.2. Transmissive type PPG

The transmissive type PPG module also uses a 3 LED configuration. Among the three LEDs, there are a red LED with a peak wavelength of 660 nm, a green LED with a peak wavelength of 530 nm, and an IR LED with a peak wavelength of 950 nm. The used LEDs are all SMD flat type. For receiving the transmitted light, the same PD from the SFH 7050 module is employed.

3.1.3. Switching and controlling unit

In both (reflective and transmissive) types, a specially designed bio-sensing analog front end (AFE) named AFE 4404 [27] is used. This

device can control three LEDs and one PD. By using an I²C interface, the analog-to-digital converter (ADC) code can be readout. Moreover, it has 24-bit resolution. For this device, a supply voltage of 2~3.6 V is necessary for the receiver, and 3~5.25 V is recommended for the transmitter. A schematic pinout diagram of the AFE4404 is shown in Fig. 3(a). In our design, 5 V is supplied to the transmitters, and 3.3 V is supplied for the PD. The Tx1, Tx2, and Tx3 pins of AFE 4404 are designated for green LED, IR LED, and red LED, respectively. The next step is to select the mode to operate between reflective and transmissive as well as to prevent functioning both modes at the same time. In order to select the mode, a solid-state relay (optocoupler MOSFET) CPC2030N [28] is used. Its pin diagram is shown in Fig. 3(b). Also, a two-input NAND gate 74AUP2G00DC-Q100H [29] is employed to prevent using both modes at the same time. Based on the output of the NAND gate, the mode is activated by the optocoupler. The pinout of the NAND gate is shown in Fig. 3(c). Another significant part of our design is the inclusion of a three-axis accelerometer. It is a generally known fact that PPG systems are affected by motion artifacts, so the MC3635 [30] serves the purpose of somehow compensating for artifacts with PPG signals. Fig. 3(d) displays the pinout diagram of the accelerometer. Fig. 4 represents the block diagram of the proposed hardware system. Overall, the tracing of the final PCB board is shown in Fig. 5 and the final commercial level PPG device is shown at different angles in Fig. 6.

3.2. Data acquisition and pre-processing

After developing the hardware system, we collect the PPG data from volunteers. For each volunteer, a total of four minutes of reflective and transmissive PPG signals are collected. At the same time, glucose levels are measured using an invasive type device named CareSens II Plus [29].

In the pre-processing stage, the PPG signal is passed through several

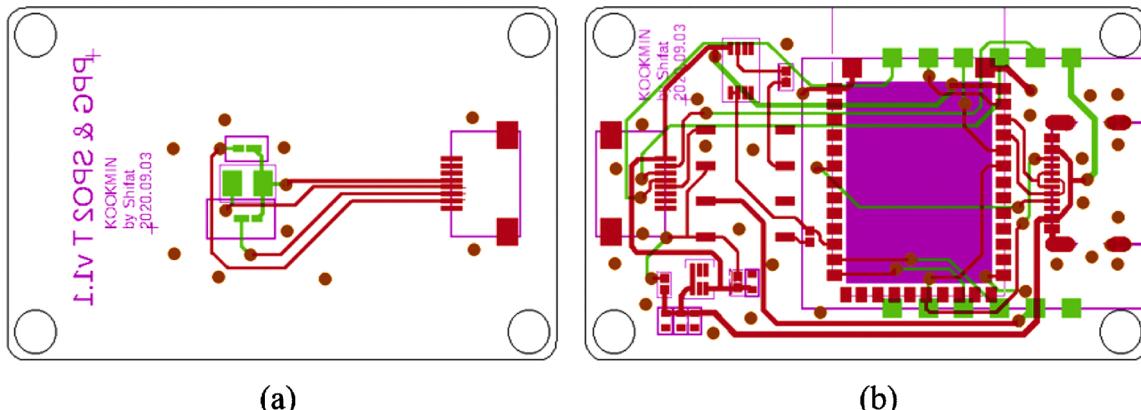


Fig. 5. Final PCB tracing of the proposed system. (a) the transmissive part, (b) the reflective part.

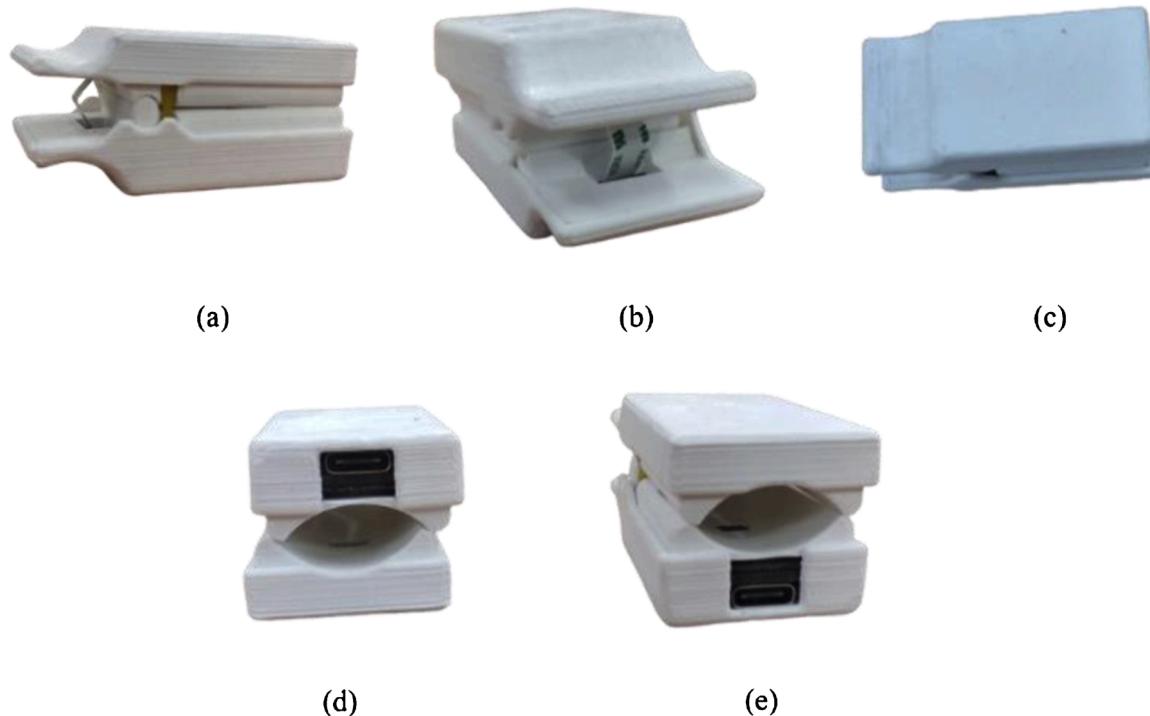


Fig. 6. Final PPG device's view from different angles. (a) Side view, (b) back view, (c) top view, (d) front view (transmissive type), and (e) front view (reflective type).

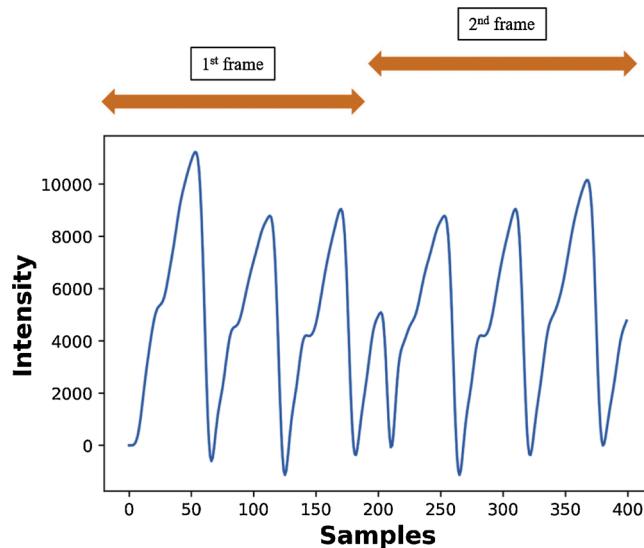


Fig. 7. Framing technique of a representative reflective mode red signal after preprocessing.

filtering steps to keep only the PPG signal part. The other high and low frequency noises are removed. Another common problem with PPG signals is baseline drifting. Therefore, we apply here fitting-based sliding window (FSW) algorithm [30]. Using this algorithm, it is convenient to detect valleys correctly. To remove the high frequency noise, a 6th order low pass Butterworth filter with -3 dB point at 8 Hz is used. Further, we segment each person's PPG signals for about 3 s as shown in Fig. 7. Then the segmented PPG signals for red and IR wavelengths are passed through feature extraction module.

3.3. Feature extraction module

In the feature extraction part, we extract seventeen significant and

discriminant features from the collected PPG signals. The features are a mixture of PPG-based physiological features, signal oriented characteristics, and physical parameter. The acquired features are as follows:

i) Zero-crossing rate (ZCR):

The number of total sign changes (from positive to negative or vice versa) in the entire signal or in a specific frame of that signal is known as ZCR. Eq. (1) deals with the mathematical expression of ZCR.

$$ZCR = \frac{1}{T-1} \sum_{t=1}^{T-1} 1_{\mathbb{R}<0}(s_t s_{t-1}) \quad (1)$$

where $1_{\mathbb{R}<0}$ denotes an indicator function, and T is the time length of signal s .

ii) Autocorrelation (ACR):

The periodical components of a signal (which is embedded in the data) can be obtained by autocorrelation [31]. Formally, it is the product of a signal and its time delayed version. The zero time-delay shows the maximum autocorrelation. The time domain version and frequency domain version of autocorrelation can be expressed by (2) and (3), respectively.

$$R_{ss}(\tau) = \int_{-\infty}^{\infty} s(t)s(t+\tau)dt \quad (2)$$

$$R_{ss}(\omega) = |S(\omega)|^2 \quad (3)$$

where τ is the time delay, and $s(t)$ and $S(\omega)$ represent the PPG signal and the corresponding Fourier transform, respectively.

iii) Power spectral density (PSD):

The distribution of energy over different frequency components of a signal can be observed by PSD. Among several methods to calculate this feature, we choose Welch's method [32] to compute PSD since this method is computationally fast enough. By Welch method, the PPG signal $s_k(m)$ is framed. For each frame

$k = 1, 2, 3, \dots, K$ then the Fourier transform $A_k(n)$ of these sequences is taken by (4). In this way, k th modified periodogram I_k is calculated using (5).

$$A_k(n) = \frac{1}{L} \sum_{m=0}^{L-1} s_k(m) W(m) e^{-2\pi j m n / L} \quad (4)$$

$$I_k(f_n) = \frac{L}{U} |A_k(n)|^2 \quad (5)$$

where

$$f_n = \frac{n}{L}, \quad U = \frac{1}{L} \sum_{m=0}^{L-1} W^2(m)$$

The PSD \hat{P} is calculated from (6) where the periodograms are averaged. The Kurtosis, variance, and mean of this PSD are then given as feature vectors.

$$\hat{P}(f_n) = \frac{1}{K} \sum_{k=1}^K I_k(f_n) \quad (6)$$

iv) Kaiser-Teager energy (KTE):

To analyze the energy profile of the acquired PPG signal while considering the signal frame as a periodic waveform, KTE [33] is used. The measurement of KTE is done according to (7). Some statistical properties such as kurtosis, skewness, variance, and mean of this energy profile are calculated.

$$\Phi[s(n)] = s^2(n) - s(n+1)s(n-1) \quad (7)$$

v) Spectral analysis:

By taking the fast Fourier transform (FFT) of the input PPG signal, the various frequency domain properties of the signal are analyzed in the study. As input features from spectral analysis, the flatness (kurtosis) and asymmetry (skewness) of the frequency distribution are considered here.

vi) Wavelet analysis:

In order to get the continuous wavelet transform (CWT) based features, the signal is modeled as "Mexican hat wavelet" as shown in (8). From the CWT, absolute mean is calculated [34].

$$\varphi(t) = \frac{2}{\sqrt{3}\sigma\pi^{1/4}} \left(1 - \left(\frac{t}{\sigma} \right)^2 \right) e^{\frac{t^2}{2\sigma^2}} \quad (8)$$

vii) Autoregressive coefficients (ARC):

The propagation of PPG signal through the blood vessel of different diameters (capillaries, veins, arteries) and the viscosity can be modeled through autoregressive (AR) model. A vector of length 5 (5th order AR model) is extracted using the Yule-Walker equation. The general representation of an AR model is like (9).

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t \quad (9)$$

where c is the constant, ϕ_i are the model parameters, p is the order, and ϵ is white noise.

viii) Blood oxygenation (SpO_2):

The percentage of oxy-hemoglobin in blood can be obtained with SpO_2 . The ratio R obtained by (10) is generally calibrated with the original value. In our case, the SpO_2 value is taken directly from the clinical device as a reference.

$$R = \frac{\left(AC/DC \right)_{\lambda_1}}{\left(AC/DC \right)_{\lambda_2}} \quad (10)$$

ix) Heart rate (HR):

The heart rate measurement from PPG signal depends on the accurate detection of peak position. In this way, the successive peak-peak distance t_{pp} is calculated and put into (11).

$$HR = \frac{60}{t_{pp}} \quad (11)$$

x) Breathing rate (BR):

The BR is calculated by using the method in [35]. By using incremental merge segmentation algorithm, three respiratory influenced parameters (amplitude, intensity, frequency) are extracted from the PPG signal. Then, these features are fused by a smart fusion algorithm to get the breathing rate.

xi) Others:

Some other important statistical features such as skewness of data distribution and sum of absolute differences are also calculated from the input PPG signal frame.

xii) Body mass index (BMI):

As BMI is closely related to the diabetes case, we include it in our feature vector. This is the only physical parameter for our proposed BGL estimation model. The mathematical equation to calculate the BMI is given by (12)

$$BMI = \frac{mass}{height^2} \quad (12)$$

So, after calculating the above mentioned features, the final feature vector can be derived as (13) for each frame f of the PPG signal s . Heart rate and breathing rate features are only used as decision functions to qualify each signal as a good PPG signal.

$$X_F^f = \begin{bmatrix} s_{zcr}, s_{ACR}, s_{PSD}^{kurt}, s_{PSD}^{var}, s_{PSD}^{mean}, s_{KTE}^{kurt}, s_{KTE}^{var}, s_{KTE}^{mean}, s_{KTE}^{skew}, s_{spec}^{kurt}, s_{spec}^{skew}, \\ s_{wavelet}^{mean}, s_{AR}^{mean}, s_{spo2}, s_{skew}, s_{sad}, BMI \end{bmatrix} \quad (13)$$

3.4. Machine learning module

Two different regression algorithms are tried at this stage. One is random forest (RF) regressor and the other is XGBoost (XGB). Both algorithms are decision tree based method. Random forest is an ensemble learning technique. The regression task in RF is operated by assembling a crowd of decision trees. On the other hand, XGB is a gradient boosting algorithm. The algorithm provides parallel tree boosting features. At the output stage, we obtain the glucose value from each regressor.

4. Results and discussions

In this section, we will show the results obtained from the hardware module as well as the final prediction of BGL by our proposed method. Moreover, a detailed discussion will be found about the pros and cons of this whole study.

4.1. Dataset description

Twenty six subjects voluntarily participate in the study. From each subject, about 4 min of PPG signal is collected. At the same time, SpO_2 is also collected by another clinical PPG device. For reference value,

Table 1
Statistics of the collected dataset.

Measurement	BMI	Age (years)	SpO ₂	BGL
Min	24.8	25	95	84
Max	34.1	80	99	199
Mean	30.31	52.4	97	119
SD	2.38	18.3	1.27	29

invasive type blood glucose is collected every time. The participants are asked to give data at pre-prandial condition. Table 1 lists the statistics of the collected dataset.

4.2. Quality of output PPG signal

In our case, we develop both reflective and transmissive type PPG systems. Fig. 8 shows the output PPG signal in the reflective part of the hardware circuit. In this figure, every column is for different wavelengths (green, red, and IR, respectively). The first row of each column is the raw signal obtained directly from the photodetector. Here, we observe some baseline drifting even though it cannot be seen in the figure because of taking only one single frame (~ 3 s). The second row shows the output after baseline wandering removal based on the FSW algorithm, and the last row is the output after Butterworth low-pass filtering. From the last row, it is seen that the high frequency noise is nicely removed from the PPG signal and the systolic and diastolic peaks even the dicrotic notch also are very clear. For transmissive mode system, we observe the signals for different wavelengths as shown in Fig. 9. From the two modes of PPG signal acquisition system, it is seen that all the signals for different wavelengths are very clear and the primary features of the PPG signals are also strongly visible except for transmissive type green LED. In Fig. 9 it can be seen that the green wavelength signal is facing a lot of noise. One of the reasons may be that the green light is scattered too often when passing through the finger

medium. Moreover, the absorbance of green light by blood is also high. Therefore, the amplitude of the signal received by the PD for the green LED is also lower compared to others. However, in the BGL estimation system, signals only from red and IR LEDs are used. Before getting into the further processing and feature extraction step, each framed signal is quantified by the measurement of heart rate (HR) and breathing rate (BR). Through this analysis, the corrupted frames are eliminated. However, only a few frames are eliminated since almost all the frames are sufficient to measure the HR and BR.

4.3. Performance of non-invasive BGL measurement

After several step data preprocessing, the PPG data comes to this step. The feature extraction module extracts several domain features and the two ML models (random forest and XGBoost) try to fit the features of the training sample to the actual BGL prediction. We follow leave-one-out method to fit the ML models between input PPG signal and BGL. In this procedure, every time one subject is used for test and others are used for training. Additionally, we include only 1% of data from each test subject to obtain a personalized ML model for individual's blood glucose prediction. The personalized training procedure is reported as an important step in [36]. The fitted models are then used to justify the proposed method upon test samples. The results shown here are only for the test subject which is considered at each iteration. The Pearson correlation coefficients (r) for different cases are tabulated in Table 2. Reflective type PPG signal performs almost similar to the transmissive system here. The BGL estimation from the reflective type has the best $r = 0.94$, while transmissive type has 0.93. Of the two regression models, XGB performs better than RF. The same performance trend can be found in Table 3, where mean absolute error (MAE) is calculated. Fig. 10 visualizes the fitted curve after regression on the testing dataset. The green line in this figure indicates the perfect fit. Almost all the predicted values are around the green line. Low and high level BGL values are also accurately predicted by our proposed method.

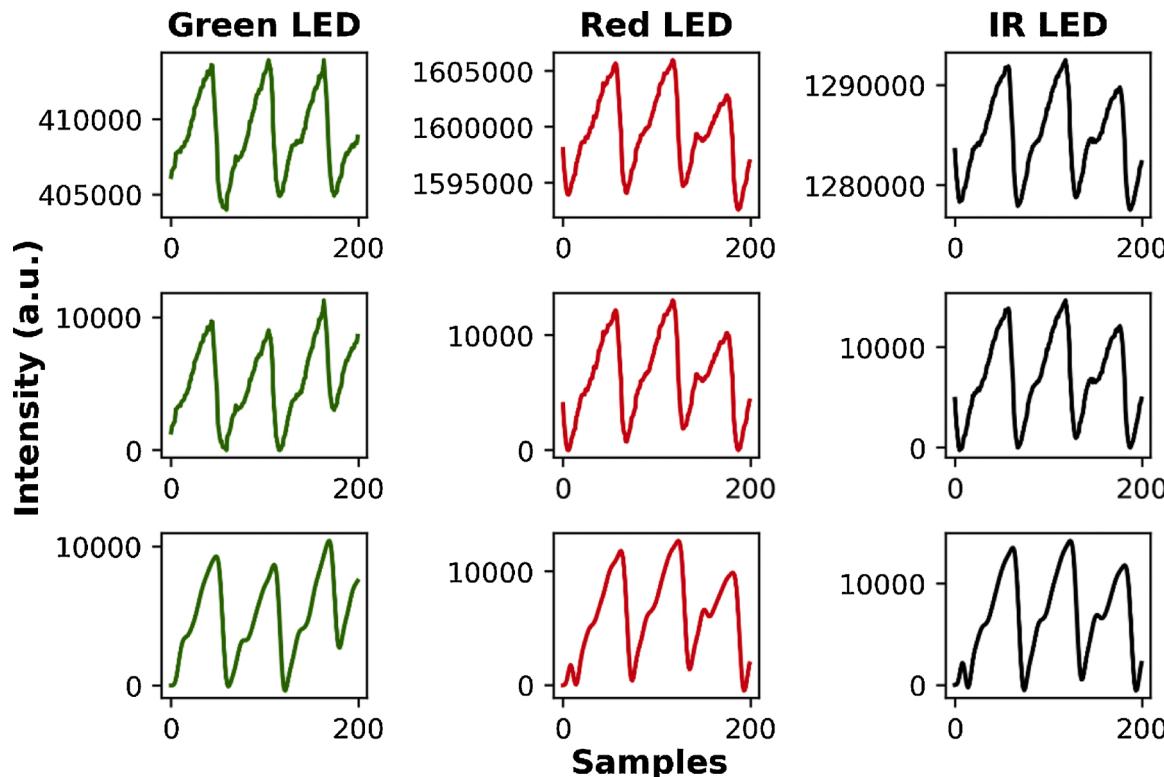


Fig. 8. Reflective type PPG signal from the developed hardware. Each column indicates different wavelength and each row indicates different pre-processing step.

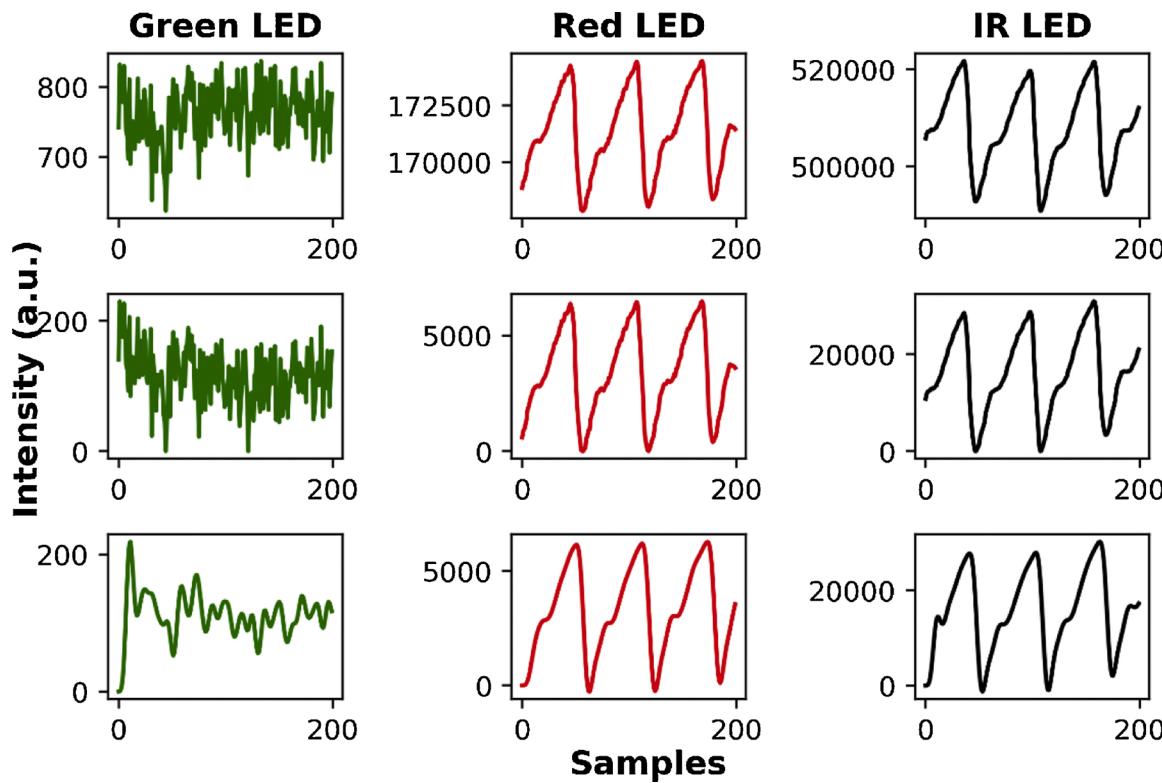


Fig. 9. Transmissive type PPG signal from the developed hardware. Each column indicates different wavelength and each row indicates different pre-processing steps.

Table 2

Pearson's correlation coefficient (r) values after regression by two methods for different cases.

ML model	PPG Data type	
	Transmissive	Reflective
Random Forest (RF)	0.88	0.88
XG Boost (XGB)	0.93	0.94

*Best values are in bold font.

Table 3

Mean absolute error (MAE) value after regression by two methods for different cases.

ML model	PPG Data type	
	Transmissive	Reflective
Random Forest (RF)	12.29	11.63
XG Boost (XGB)	8.73	8.31

*Best values are in bold font.

To verify the clinical safety of our proposed non-invasive BGL estimation method, the mostly used Clarke-Error grid analysis (EGA) [37] is done here. The whole grid is divided into five regions (A, B, C, D, E). Among these areas, region A shows the values which are within 20% of ground truth values and these values are clinically perfect. Region B indicates the values which are beyond 20% from reference values but still clinically usable (safe). Region C encircles the points leading to unnecessary treatment. The most critical zones are D and E. While values in D region indicate the failure to detect hyperglycemia or hypoglycemia, points based on area E create severe confusion in treatment. In our situation, we can see from Figs. 11 and 12 that the predicted BGL values are within regions A and B in both cases (reflective and transmissive

PPG). Table 4 lists the zone percentage of the predicted values. The analysis shows that no values are falling into dangerous regions C, D, and E. Considering the percentage of points that fall in the safe zones A and B in the Clarke error grid plot, we see that, for transmissive type, RF performs better than XGB. On the other hand, for reflective type, the XGB algorithm performs better than the RF. Our algorithm can almost accurately predict glucose levels between 84~199 mg/dL, however, the algorithm makes some errors around 100 and 150 mg/dL. This trend can be seen in both Figs. 11 and 12. In Fig. 11(a) and (b), some points come to lower and upper B region of the EGA plot. This is because the algorithm somehow predicts the glucose value about 150 mg/dL and 100 mg/dL beyond 20% lower and higher, respectively. Overall, the performance of the proposed BGL estimation model is satisfactory since it can predict very accurately from moderately low to higher levels of glucose. To test these statements on the EGA plot for our method, we can now look at the Fig. 13, which is called Bland-Altman (B&A) analysis. In this plot, the green line represents the mean difference (\bar{d}) between the two measurements (predicted glucose value by our method and the glucose value obtained by invasive device), and the red lines represent $\bar{d} \pm 1.96 \times std$ where std stands for standard deviation. First, focusing on the mean difference between the two measurements, we can see that \bar{d} is very close to zero value in all cases, which is very good. Also, most of the values fall around mean difference line and within the level of agreements (red lines). So, apart from the correlation coefficient, we can say that the predicted BGL values by our proposed method are also reliable as a result of analyzing the EGA and B&A plots.

Also, we compared our proposed method with the other two established glucose estimation methods. The method in [9] estimated glucose by using machine learning techniques and took reflective data from red and infrared light. In [17], a deep CNN method was used on 1D raw PPG signals. Our method achieved almost same Pearson's r as the method of [17] as shown in Table 5.

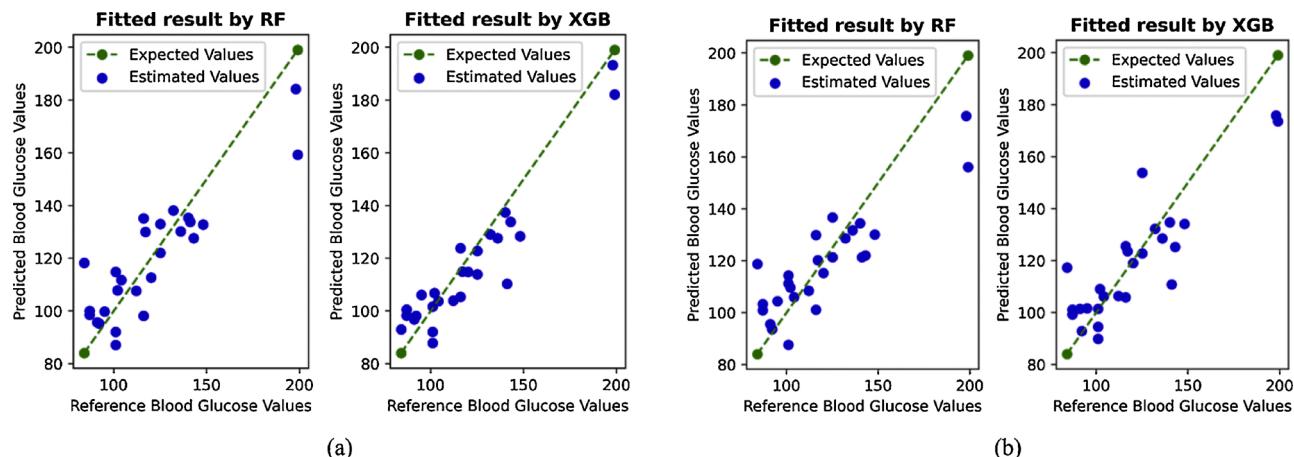


Fig. 10. Fitted scatter plots by different regressors (a) for reflective type, and (b) for transmissive type.

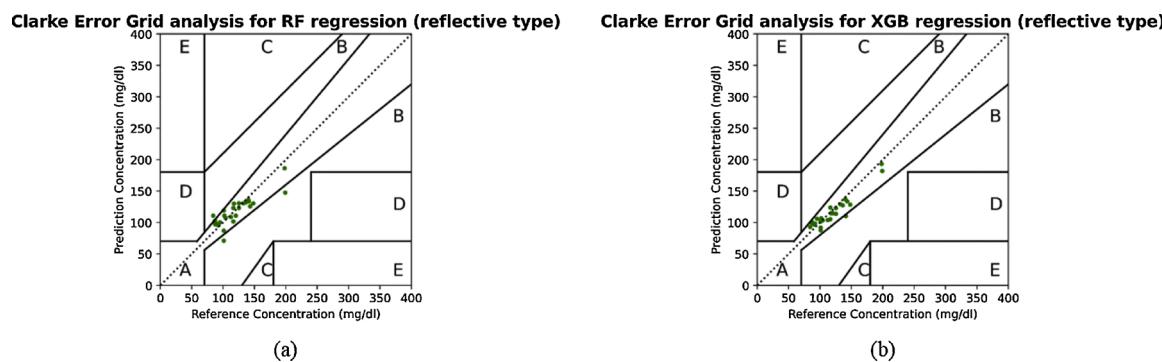


Fig. 11. EGA plots for BGL estimation using reflective type PPG signal. (a) using RF regressor, (b) using XGB regressor.

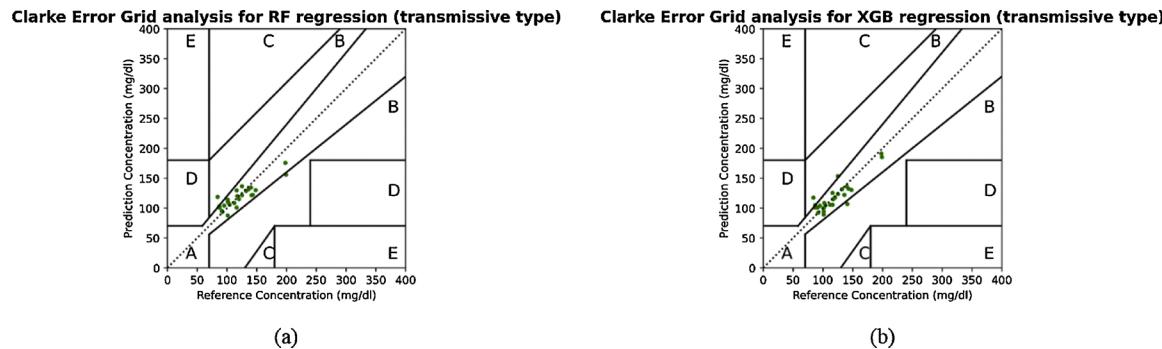


Fig. 12. EGA plots for BGL estimation using transmissive type PPG signal. (a) using RF regressor, (b) using XGB regressor.

Table 4
Zone accuracy of EGA plots.

Type	Zone				
	A	B	C	D	E
Transmissive (RF)	92.30%	7.70%	0%	0%	0%
Transmissive (XGB)	85%	15%	0%	0%	0%
Reflective (RF)	88.47%	11.53%	0%	0%	0%
Reflective (XGB)	96.15%	3.85%	0%	0%	0%

*Best values are in bold font.

4.4. Limitations and future research

In the hardware side, one of the limitations of this research is that we do not get a clean transmissive PPG signal when green light passes

through the body medium. However, this problem will be solved after replacing it with a newly released chip. Moreover, the accelerometer data is not utilized yet. We hope this will help to detect and eliminate motion artifacts from the signal in the future. In the processing site, we see that the ML algorithms find the BMI feature very important. The feature importance plot can be seen in the Appendix A. However, only the BMI feature is not sufficient to predict BGL non-invasively (very low Pearson's r value). Therefore, our future study will try to estimate the BGL level without the person's physical parameter BMI. Moreover our effort will also be in collecting more diverse and larger dataset including hypoglycemic and hyperglycemic subjects' BGL data. However, in this given short dataset our method can detect certain BGL levels (84–199 mg/dL) effectively. Despite having some limitations, the proposed hardware and processing system is suitable as a proof of study

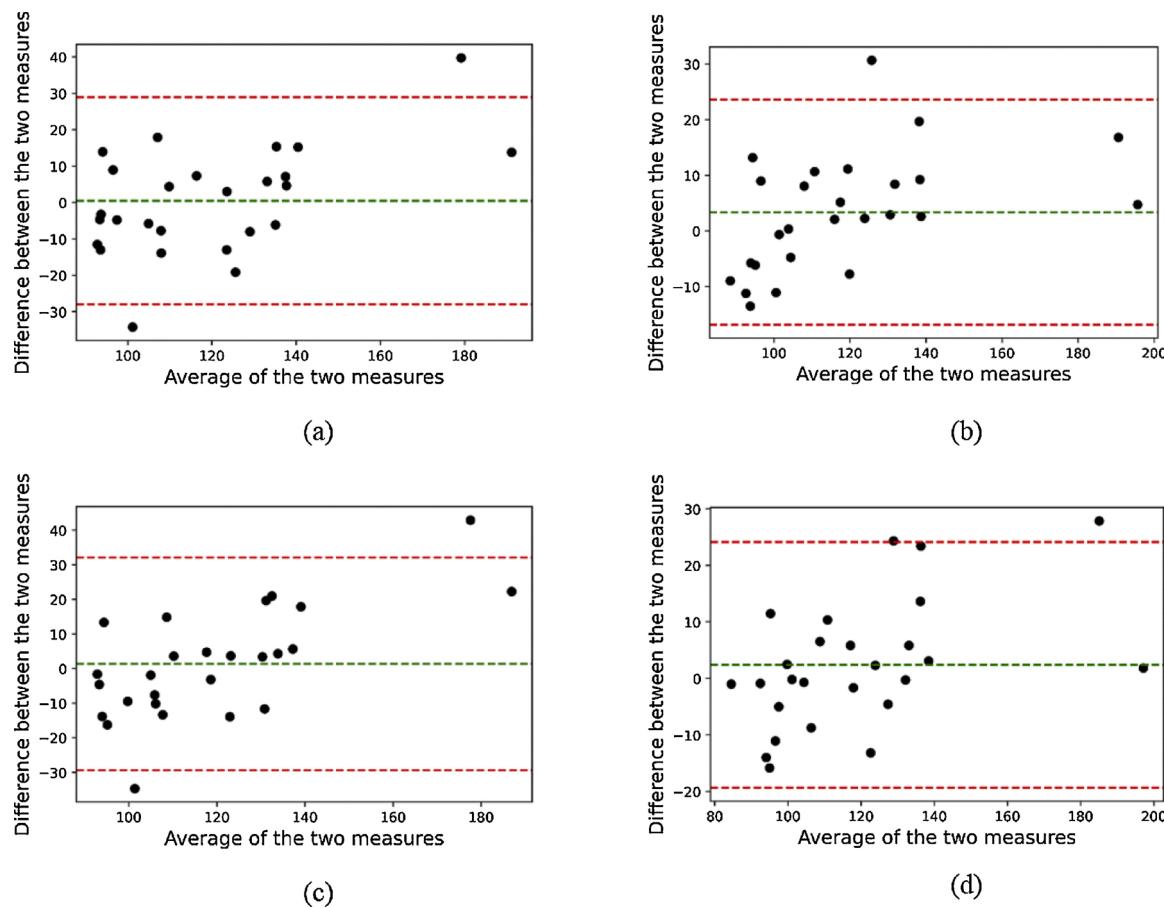


Fig. 13. B&A plots for predicted values and reference values. (a) reflective PPG and RF regressor, (b) reflective PPG and XGB regressor, (c) transmissive PPG and RF regressor, and (d) transmissive PPG and XGB regressor.

Table 5
Comparison of Pearson's r with other methods.

Ours	Method [9]	Method [17]
0.94	0.90	0.95

*Best values are in bold font.

purpose.

5. Conclusion

This research is primarily devoted to developing a commercial level compact PPG signal acquisition system. The developed PPG system comprises both reflective and transmissive systems. As the system contains red, green, and IR LEDs, which are mostly used to illuminate the skin in the field of PPG, it is possible to measure basic health parameters such as HR, SpO₂ from our proposed device. In the analysis of the PPG signal quality, we have also found that in all cases the output signal of the system is suitable for measuring the body parameters. In the second focus, we tried to measure BGL very accurately using machine learning on a very small number of features. We found favorable results in both statistical and clinical plots. Among several different ways, reflective type signals with XGB regressor performed best quantitatively. In fact, as we discussed, some limitations still exist. In the future, there will be an attempt to overcome these and make the device wearable.

Declaration

This statement is to certify that all Authors have seen and approved

the manuscript being submitted. We warrant that the article is the Authors' original work and has not received prior publication and is not under consideration for publication elsewhere.

CRediT authorship contribution statement

Shantanu Sen Gupta: Conceptualization, Methodology, Software, Writing-Original draft preparation. **Tae-Ho Kwon:** Software, Validation, Formal analysis. **Shifat Hossain:** Hardware, Resources. **Ki-Doo Kim:** Validation, Supervision, Project administration, Funding acquisition.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Education (NRF-2019R1F1A1062317) and was also supported by the National Research Foundation of Korea Grant funded by the Ministry of Science, ICT, and Future Planning (2015R1A5A7037615).

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A

The feature importance plots for different regressors (RF and XGB) in case of different mode (transmissive and reflective) are shown in Fig. A1.

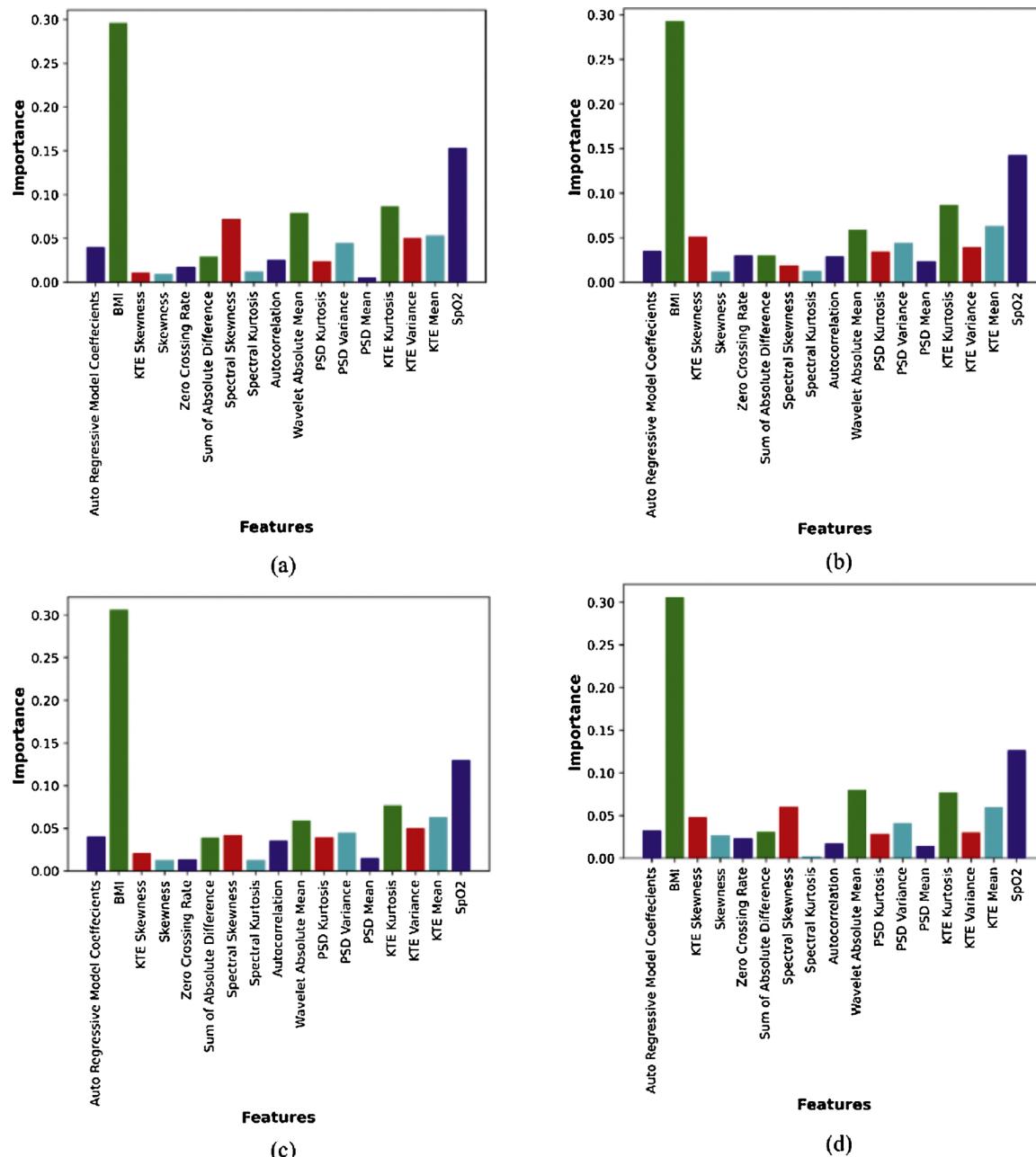


Fig. A1. Feature importance plots for (a) reflective PPG and RF regressor, (b) reflective PPG and XGB regressor, (c) transmissive PPG and RF regressor, and (d) transmissive PPG and XGB regressor.

Although, the BMI feature is given more importance, only BMI without other features shows very poor performance (very low Pearson's r).

References

- [1] I. Fine, The optical origin of the PPG signal, in: Saratov Fall Meeting 2013: Optical Technologies in Biophysics and Medicine XV; and Laser Physics and Photonics XV, vol. 9031, 2014, <https://doi.org/10.1117/12.2051228>, p. 9031.3 Jan.
- [2] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, H. Nazeran, A review on wearable photoplethysmography sensors and their potential future applications in health care, Int. J. Biosens. Bioelectron. 4 (4) (2018) 195–202, <https://doi.org/10.15406/ijbsbe.2018.04.00125>.
- [3] Y. Athavale, S. Krishnan, Biosignal monitoring using wearables: observations and opportunities, Biomed. Signal Process. Control 38 (September) (2017) 22–33, <https://doi.org/10.1016/j.bspc.2017.03.011>.
- [4] M. Poh, N.C. Swenson, R.W. Picard, Motion-tolerant magnetic earring sensor and wireless earpiece for wearable photoplethysmography, IEEE Trans. Inf. Technol. Biomed. 14 (May (3)) (2010) 786–794, <https://doi.org/10.1109/TITB.2010.2042607>.
- [5] Y.K. Lee, J. Jo, H.S. Shin, Development and evaluation of a wristwatch-type photoplethysmography array sensor module, IEEE Sens. J. 13 (May (5)) (2013) 1459–1463, <https://doi.org/10.1109/JSEN.2012.2235424>.
- [6] H. Lee, H. Ko, C. Jeong, J. Lee, Wearable photoplethysmographic sensor based on different LED light intensities, IEEE Sens. J. 17 (February (3)) (2017) 587–588, <https://doi.org/10.1109/JSEN.2016.2633575>.
- [7] E.M.G. Rodrigues, R. Godina, C.M.P. Cabrita, J.P.S. Catalão, Experimental low cost reflective type oximeter for wearable health systems, Biomed. Signal Process. Control 31 (January) (2017) 419–433, <https://doi.org/10.1016/j.bspc.2016.09.013>.
- [8] P.P. Banik, S. Hossain, T.-H. Kwon, H. Kim, K.-D. Kim, Development of a wearable reflection-type pulse oximeter system to acquire clean PPG signals and measure puls rate and SpO2 with and without finger motion, Electronics 9 (November (11)) (2020), 11, <https://doi.org/10.3390/electronics9111905>.
- [9] E. Monte-Moreno, Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques, Artif. Intell. Med. 53 (October (2)) (2011) 127–138, <https://doi.org/10.1016/j.artmed.2011.05.001>.

- [10] E.M. Moreno, et al., Type 2 diabetes screening test by means of a pulse oximeter, *IEEE Trans. Biomed. Eng.* 64 (February (2)) (2017) 341–351, <https://doi.org/10.1109/TBME.2016.2554661>.
- [11] A. Hina, W. Saadeh, A noninvasive glucose monitoring SoC based on single wavelength photoplethysmography, *IEEE Trans. Biomed. Circuits Syst.* 14 (June (3)) (2020) 504–515, <https://doi.org/10.1109/TBCAS.2020.2979514>.
- [12] X. Zhou, B.-W.-K. Ling, Z. Tian, Y.-W. Ho, K.-L. Teo, Joint empirical mode decomposition, exponential function estimation and L 1 norm approach for estimating mean value of photoplethysmogram and blood glucose level, *IET Signal Process.* 14 (September (9)) (2020) 652–665, <https://doi.org/10.1049/iet-spr.2020.0096>.
- [13] V. Dantu, J. Vempati, S. Srivilliputhur, Non-invasive blood glucose monitor based on spectroscopy using a smartphone, in: *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2014, 2014, pp. 3695–3698, <https://doi.org/10.1109/EMBC.2014.6944425>.
- [14] B. Paul, M.P. Manuel, Z.C. Alex, Design and development of non invasive glucose measurement system, 2012 1st International Symposium on Physics and Technology of Sensors (ISPTS-1) (2012) 43–46, <https://doi.org/10.1109/ISPTS.2012.6260873>. March.
- [15] W. Liu, A. Huang, P. Wang, BpMC: a novel algorithm retrieving multilayered tissue bio-optical properties for non-invasive blood glucose measurement, 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (2017) 451–456, <https://doi.org/10.1109/BIBM.2017.8217690>. November.
- [16] S. Ramasahayam, L. Arora, S.R. Chowdhury, M. Anumukonda, FPGA based system for blood glucose sensing using photoplethysmography and online motion artifact correction using adaline, 2015 9th International Conference on Sensing Technology (ICST) (2015) 22–27, <https://doi.org/10.1109/ICST.2015.7438358>. December.
- [17] S. Hossain, B. Debnath, S. Biswas, Md. J. Al-Hossain, A. Anika, S.K. Zaman Navid, Estimation of blood glucose from PPG Signal using convolutional neural network, 2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON) (2019) 53–58, <https://doi.org/10.1109/BECITHCON48839.2019.9063187>. November.
- [18] Md.M. Islam, S.M. Manjur, Design and implementation of a wearable system for Non-invasive glucose level monitoring, 2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON) (2019) 29–32, <https://doi.org/10.1109/BECITHCON48839.2019.9063193>. November.
- [19] R. Avram, et al., A digital biomarker of diabetes from smartphone-based vascular signals, *Nat. Med.* (August) (2020) 1–7, <https://doi.org/10.1038/s41591-020-1010-5>.
- [20] Y. Zhang, J.-M. Zhu, Y.-B. Liang, H.-B. Chen, S.-M. Yin, Z.-C. Chen, Non-invasive blood glucose detection system based on conservation of energy method, *Physiol. Meas.* 38 (2) (2017) 325–342, <https://doi.org/10.1088/1361-6579/aa50cf>.
- [21] S. Habbu, M. Dale, R. Ghongade, Estimation of blood glucose by non-invasive method using photoplethysmography, *Sādhanā* 44 (May (6)) (2019) 135, <https://doi.org/10.1007/s12046-019-1118-9>.
- [22] A. Hina, H. Nadeem, W. Saadeh, A single LED photoplethysmography-based noninvasive glucose monitoring prototype system, 2019 IEEE International Symposium on Circuits and Systems (ISCAS) (2019) 1–5, <https://doi.org/10.1109/ISCAS.2019.8702747>. May.
- [23] T.T. Chowdhury, T. Mishma, S. Osman, T. Rahman, Estimation of blood glucose level of type-2 diabetes patients using smartphone video through PCA-DA, in: *Proceedings of the 6th International Conference on Networking, Systems and Security*, New York, NY, USA, Dec, 2019, pp. 104–108, <https://doi.org/10.1145/3362966.3362983>.
- [24] C.-M. Li, et al., Synchronizing chaoticification with support vector machine and wolf pack search algorithm for estimation of peripheral vascular occlusion in diabetes mellitus, *Biomed. Signal Process. Control* 9 (January) (2014) 45–55, <https://doi.org/10.1016/j.bspc.2013.10.001>.
- [25] ESP32-PICO-V3 Espressif Systems | Mouser, Mouser Electronics. <https://kr.mouser.com/ProductDetail/356-ESP32-PICO-V3/> (Accessed 03 January 2021).
- [26] “BIOFY®, SFH 7050 | OSRAM Opto Semiconductors.” https://www.osram.com/ecat/BIOFY%C2%AE%20SFH%207050/com/en/class_pim_web_catalog_103489/prd_pim_device_2220012/ (Accessed 03 January 2021).
- [27] “AFE4404 data sheet, product information and support | TI.com.” <https://www.ti.com/product/AFE4404#product-details##params> (Accessed 03 January 2021).
- [28] CPC2030N IXYS Integrated Circuits | Mouser, Mouser Electronics. <https://kr.mouser.com/ProductDetail/849-CPC2030N/> (Accessed 06 January, 2021).
- [29] “Welcome to CareSense.” http://caresens.co.kr/01_product/051_product.php (Accessed 07 January, 2021).
- [30] G. Zhang, et al., A noninvasive blood glucose monitoring system based on smartphone PPG signal processing and machine learning, *IEEE Trans. Ind. Inform.* 16 (November (11)) (2020) 7209–7218, <https://doi.org/10.1109/TII.2020.2975222>.
- [31] *Acquisition and Processing of Marine Seismic Data*, Elsevier, 2018.
- [32] P. Welch, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, *IEEE Trans. Audio Electroacoustics* 15 (June (2)) (1967) 70–73, <https://doi.org/10.1109/TAU.1967.1161901>.
- [33] J.F. Kaiser, Some useful properties of teager's energy operators, in: 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, 1993, pp. 149–152, <https://doi.org/10.1109/ICASSP.1993.319457>. Apr.
- [34] S. Mallat, *A Wavelet Tour of Signal Processing*, Third, Elsevier, 2009.
- [35] W. Karlen, S. Raman, J.M. Anserrmino, G.A. Dumont, Multiparameter respiratory rate estimation from the photoplethysmogram, *IEEE Trans. Biomed. Eng.* 60 (July (7)) (2013) 1946–1953, <https://doi.org/10.1109/TBME.2013.2246160>.
- [36] G. Slapničar, N. Mlakar, M. Luštrek, Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network, *Sensors* 19 (January (15)) (2019), 15, <https://doi.org/10.3390/s19153420>.
- [37] W.L. Clarke, D. Cox, L.A. Gonder-Frederick, W. Carter, S.L. Pohl, Evaluating clinical accuracy of systems for self-monitoring of blood glucose, *Diabetes Care* 10 (September (5)) (1987) 622–628, <https://doi.org/10.2337/diacare.10.5.622>.