https://doi.org/10.7840/kics.2023.48.7.852

Noninvasive In-Vivo Estimation of Blood-Glucose Concentration Using Beer-Lambert-Based Model

Chowdhury Azimul Haque◆, Shifat Hossain*, Tae-Ho Kwon**,
Hyoungkeun Kim***, Ki-Doo Kim°

ABSTRACT

Noninvasive measurement of blood-glucose concentration can reduce both pain and complications associated with piercing the human fingertip to collect blood. Photoplethysmography (PPG) is a helpful technique that can be used to measure blood-glucose concentration without a blood sample. To facilitate such noninvasive in-vivo estimation, we propose a model based on the Beer-Lambert law for measuring blood-glucose concentration using the PPG signal. Notably, only two wavelengths are used. First, the oxygen saturation (SpO2) is estimated from the ratio of absorbance at two wavelengths, then another absorbance ratio is presented, and the blood-glucose concentration is estimated by substituting the SpO2 estimated earlier to this ratio. The PPG signals from 40 subjects were collected along with their reference blood-glucose concentrations and SpO2 values. The PPG-based blood-glucose concentrations are then calculated using mathematical equations derived from the Beer-Lambert law. A supervised machine learning model, XGBoost, is applied to calibrate the estimation model with the reference values measured using a commercial device; according to our experimental results, the Pearson correlation coefficient (Pearson's r) value is 0.85. The proposed model based on the Beer-Lambert law thus provides a method for in-vivo estimation of blood glucose in daily applications.

Key Words : Blood-glucose concentration; In-vivo estimation; Noninvasive measurement; Beer-Lambert law; Photoplethysmography

Ⅰ. Introduction

The blood glucose concentration in the human body typically represents the amount of glucose present in the blood. The human body regulates blood glucose concentrations with the goal of maintaining adequate levels, that is, sufficient glucose to fuel the cells but not excessive enough to place an undue burden on the blood circulation system. Diabetes mellitus (DM) is a serious disease affecting more than 422 million people worldwide and causes high blood glucose levels when not properly treated. According to the American Diabetes Association[1], based on the blood-glucose concentration present in the human body, health status can be categorized into three types: normal, pre-diabetes, and diabetes. The blood glucose

※ This work was supported by the National Research Foundation (NRF) of Korea funded by the Ministry of Science and ICT (2022R1A5A7000765) and by the Basic Science Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Education (NRF-2022R1A2C2010298). This research was supported by the Korea Industrial Technology Association (KOITA), funded by the Ministry of Science and ICT (MSIT).
◆ First Author : Kookmin University Department of Electronic Engineering, c_azimul@kookmin.ac.kr, 학생회원
° Corresponding Author : Kookmin University Department of Electronic Engineering, kdk@kookmin.ac.kr, 종신회원
* Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL 32816, USA; shifathosn@knights.ucf.edu, 학생회원
** Department of Electronics Engineering, Kookmin University, Seoul 02707, Korea, 정회원
*** Korea I.T.S. Co., Ltd., Gangnam, Seoul 06373, Korea
논문번호 : 202303-051-C-RU, Received March 15, 2023; Revised April 18, 2023; Accepted April 19, 2023

concentration test can be performed in two states: fasting state (fasting means not having anything to eat or drink except water for at least 8 h before the test) and post-meal state (blood-glucose concentration two hours after meals). In the fasting state, blood glucose concentrations less than 100 mg/dL are considered normal. If it ranges from 100 mg/dL to 125 mg/dL, it is considered prediabetic. For diabetes, the blood glucose concentration was > 126 mg/dL. Therefore, the continuous measurement of blood glucose concentration is an essential approach that allows people to plan healthy lifestyles. However, invasive blood glucose concentration measurements not only aggravate the pain of individuals but also create a burden on them[2]. Nevertheless, noninvasive blood glucose testing frameworks for self-monitoring are still in the early stages of improvement and are still far from being purposed for at-home usage.

Photoplethysmography (PPG) signals are used to estimate blood flow in the skin using light. Traditionally, oxygen saturation (SpO2), blood pressure, and cardiac output have been used to assess autonomic functions. Over the past few decades, PPG signals have been applied in various clinical assessments, like monitoring blood oxygen saturation, blood pressure, heart rate variability, and glucose levels. PPG can be used to estimate blood pressure using biometric parameters such as body mass index (BMI), height, and age[3]. The heart rate is an important variable for measuring a wide range of physiological parameters. The AC component of the PPG signal can be coordinated with the pulsating heart rate and becomes a source of heart rate information[4]. Quantitative parameters, like blood glucose concentration, can also be detected from PPG signals. Monte Moreno et al.[5] used various physiological parameters such as heart rate, vascular compliance, blood viscosity, and respiratory frequency to analyze PPG waveforms and subsequently estimate blood glucose. Support vector machine (SVM), random forest, linear regression, and neural network classifiers have been used for glucose classification, among which random forest performs the best, with an R2 value of 0.9. In [6], a noninvasive blood glucose monitoring system based on PPG signal processing and a machine-learning algorithm was proposed. PPG signals were extracted from videos of the index finger. To remove baseline drift, a four-sliding-window (FSW) pattern-matching algorithm was developed, and 28 time-domain and frequency-domain features were extracted using the Gaussian fitting method. Eventually, the blood glucose concentrations were classified into three groups—normal, borderline, and warning levels—using a machine learning algorithm, such that the 28 extracted features were effective in separating the glucose concentrations. Hossain et al.[7] developed a system based on a convolutional neural network (CNN) to measure blood glucose noninvasively using PPG signals and created a dataset of PPG signals and blood glucose concentrations from 30 subjects. The Pearson correlation coefficient of their regression analysis was 0.95 for estimating glucose concentration. Tsai et al.[8] utilized Glutrac, a health device, to obtain PPG signals from the fingers and wrists, which were then sent to a NoSQL database. After filtering the signals, 30 time-domain features were extracted, and the first derivative was obtained. Multiple decision trees were used in the data learning investigation, and the accuracy of the model was 80%. Furthermore, Gupta et al.[9] developed a device to record PPG signals in both transmission and reflectance modes. Seventeen features were extracted from the filtered signals and used to establish their relationship with blood glucose concentration. The maximum Pearson's correlation coefficient for this methodology was 0.94. In [10], PPG signals were collected at different wavelengths using wearable biosensors and the baseline drifts of the signals were removed using a digital wavelet transform along with wavelet decomposition. Then, 24 features were extracted from each periodic signal by applying the local maxima algorithm for peak signal detection; a total of 12 subjects participated in this experiment. Ten-fold cross-validation was used to evaluate the linear partial least-squares, multivariate-based calibration model. The correlation coefficient for this experiment was 0.86. Recently, wearable devices have emerged as a promising method for collecting PPG signals and using them for further processing. Banik et al.[11] developed a wearable device to collect PPG

signals and measure the heart rate and SpO2 values. The authors in [12] derived a mathematical model based on the Beer-Lambert law to estimate HbA1c using digital volume pulse signals, that is PPG. Moreover, in our recent study[13], we estimated the blood glucose concentration noninvasively using Monte Carlo simulation (MCS), wherein MCS-based photon propagation in a finger model was used to estimate the blood glucose concentration.

Beer-Lambert law defines the linear relationship between the absorbance and molar concentration, molar absorption coefficient, and the optical path of a substance or solution. According to this law, if light passes through a substance or solution, then the absorbance can be defined using the molar absorption coefficient, molar concentration, and optical path length of the light that traverses the solution. There are two types of PPG signals based on the placement of the light source and the PD. If the light source and PD are placed on the same side, they are considered to be reflectance-type PPG signals, and if the light source and PD are placed on opposite sides, they are considered to be transmission-type signals[14]. Therefore, for a transmission-type PPG signal, if the signal measurement site is considered to be a substance, then the absorbance of the substance can be defined by the Beer-Lambert law.

To design a system to measure blood glucose concentrations as a point-of-care monitoring tool, this study proposes a Beer-Lambert-law-based model for noninvasive in vivo estimation of blood glucose concentration using PPG signals. The PPG signals were recorded from 40 volunteers at two LED wavelengths: 950 nm (infrared) and 660 nm (red). In addition to the PPG signals of the subjects, their SpO2 and blood glucose concentrations were measured as references using typical clinical devices. The raw PPG signals were filtered appropriately, and the filtered signals were used to calculate the ratios for estimating blood glucose concentrations. Thereafter, a supervised machine-learning algorithm, XGBoost, was used to calibrate the estimated concentration using the reference, and the model performance was listed accordingly. The application of our current study can alleviate the patient burden of checking blood glucose concentration with finger pricking upon the use of this method on a daily basis.

The remainder of this paper is organized as follows: the methodology, including derivation of the mathematical equations, data acquisition, preprocessing stages, and calibration, is described in Section 2. In Section 3, the results of the blood glucose concentration estimations are described. Finally, a discussion of this work and the conclusions are presented in Sections 4 and 5, respectively.

Ⅱ. Methodology

The workflow diagram of this study is as shown in Fig. 1.

2.1 Beer-Lambert-Law-Based Model

The Beer-Lambert law specifies the attenuation of light through the sample it travels. In most cases, the Beer-Lambert law is suitable for quantifying the concentrations of compounds remaining in the

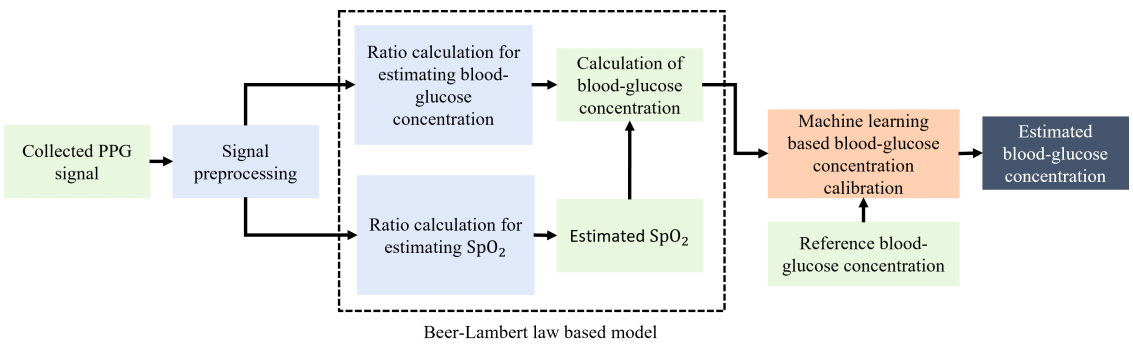


Fig. 1. Workflow diagram of this study.

samples. Accordingly, light attenuation is directly proportional to the concentration of residual compounds in the sample[15]. The practical expression of the Beer-Lambert law is given by (1).

$$A = \varepsilon \times c \times d \tag{1}$$

where A is the total absorption, ε is is the molar attenuation coefficient or absorptivity of the attenuating species ($L \cdot mol^{-1} \cdot cm^{-1}$), c is the concentration of the attenuating species ($mol \cdot cm^{-1}$), and d is the optical path length (cm). Equation (1) can also be expressed in terms of the incident light intensity of the sample and light intensity transmitted through the sample.

$$A = \log \frac{I_0}{I} \tag{2}$$

where I_0 is the intensity of the incident light on the sample and I is the intensity of the light transmitted through the sample.

Considering equation (1), blood can be defined as a homogeneous solution of hemoglobin and glucose. Hence, the total absorption coefficient at wavelength λ can be formulated as:

$$A = (\varepsilon_{HbO}(\lambda) \times c_{HbO} + \varepsilon_{HHb}(\lambda) \times c_{HHb} + \varepsilon_G(\lambda) \times c_G) \times d \tag{3}$$

where $\varepsilon_{HbO}(\lambda)$, $\varepsilon_{HHb}(\lambda)$, *and* $\varepsilon_G(\lambda)$ are the molar absorption coefficients at wavelength λ for HbO (oxygenated hemoglobin), HHb (deoxygenated hemoglobin), and glucose, respectively; c represents the molar concentration of each element, and d is the distance traveled by light.

In our model, blood is represented as a homogeneous solution of hemoglobin consisting of oxygenated hemoglobin (HbO), deoxygenated hemoglobin (HHb), and glucose. Hence, the%SpO_2 can be described as follows:

$$\%SpO_2 = \frac{c_{HbO}}{c_{HbO} + c_{HHb}} \times 100\% \tag{4}$$

where c_{HbO} and c_{HHb} represent the molar concentrations of HbO and HHb, respectively. Blood oxygen saturation (SpO_2) is related to oxygen-bonded hemoglobin[16], but other substances present in the blood are not. Thus, the right-hand side of (4) did not contain any substances other than HbO or HHb.

The partial molar concentrations of HbO, HHb, and glucose are expressed as P_{HbO}, P_{HHb}, and P_G, respectively, as follows:

$$P_{HbO} = \frac{c_{HbO}}{c_{HbO} + c_{HHb} + c_G} \tag{5}$$

$$P_{HHb} = \frac{c_{HHb}}{c_{HbO} + c_{HHb} + c_G} \tag{6}$$

$$P_G = \frac{c_G}{c_{HbO} + c_{HHb} + c_G} \tag{7}$$

$$c_{HbO} + c_{HHb} + c_G = c_T \tag{8}$$

From (5) - (7),

$$P_{HHb} = 1 - (P_G + P_{HbO}) \tag{9}$$

Using the partial molar concentrations (P_{HbO} and P_{HHb}), (4) can be expressed as

$$\%SpO_2 = \frac{P_{HbO}}{P_{HHb} + P_{HbO}} \times 100\% \tag{10}$$

According to the blood vessel model explained in [12], the diameter of the vessel expands when blood enters the vessel (systolic phase) and decreases when blood flows out from the vessel (diastolic phase). Considering this model, the difference between the two states (i.e., when blood enters and flows out of the vessel) of light traversal in the finger can be obtained in the form of the Beer-Lambert law, as shown in (11).

$$\Delta A = (\varepsilon_{HbO}(\lambda) \times c_{HbO} + \varepsilon_{HHb}(\lambda) \times c_{HHb} + \varepsilon_G(\lambda) \times c_G) \times \Delta d \tag{11}$$

where ΔA = A1 - A2, $\Delta d = d1 - d2$; $A1$ represents the absorbance when blood enters the vessel, and $A2$ represents the absorbance when the blood flows out from the vessel. Variables $d1$ and $d2$ represent the diameters of the blood as it enters and leaves the vessel, respectively.

By replacing c_{HbO}, c_{HHb} and c_G with their partial concentration values, and eliminating P_{HHb} using (9), we can rewrite (11) as follows:

$$\Delta A = c_T(\varepsilon_G(\lambda) \times P_G - \varepsilon_{HHb}(\lambda) \times P_G - \varepsilon_{HHb}(\lambda) \times P_{HbO} + \varepsilon_{HHb}(\lambda) + \varepsilon_{HbO}(\lambda) \times P_{HbO}) \times \Delta d \quad (12)$$

For the two wavelengths considered in this study (i.e., $\lambda 1$= 950 nm and $\lambda 2$ = 660 nm), (12) can be expressed as

$$\Delta A_{\lambda 1} = c_T(\varepsilon_G(\lambda 1) \times P_G - \varepsilon_{HHb}(\lambda 1) \times P_G - \varepsilon_{HHb}(\lambda 1) \times P_{HbO} + \varepsilon_{HHb}(\lambda 1) + \varepsilon_{HbO}(\lambda 1) \times P_{HbO}) \times \Delta d \quad (13)$$

$$\Delta A_{\lambda 2} = c_T(\varepsilon_G(\lambda 2) \times P_G - \varepsilon_{HHb}(\lambda 2) \times P_G - \varepsilon_{HHb}(\lambda 2) \times P_{HbO} + \varepsilon_{HHb}(\lambda 2) + \varepsilon_{HbO}(\lambda 2) \times P_{HbO}) \times \Delta d \quad (14)$$

From (13) and (14), a ratio equation can be obtained and used to estimate unknown parameter P_G. The ratio equation is expressed as

$$R = \frac{\Delta A_{\lambda 2}}{\Delta A_{\lambda 1}} = \frac{(\varepsilon_G(\lambda 2) \times P_G - \varepsilon_{HHb}(\lambda 2) \times P_G - \varepsilon_{HHb}(\lambda 2) \times P_{HbO} + \varepsilon_{HHb}(\lambda 2) + \varepsilon_{HbO}(\lambda 2) \times P_{HbO})}{(\varepsilon_G(\lambda 1) \times P_G - \varepsilon_{HHb}(\lambda 1) \times P_G - \varepsilon_{HHb}(\lambda 1) \times P_{HbO} + \varepsilon_{HHb}(\lambda 1) + \varepsilon_{HbO}(\lambda 1) \times P_{HbO})} \quad (15)$$

Equations (13) and (14) can be expressed in the form of (2) as follows:

$$\Delta A_{\lambda 1} = \Delta \left[\log \frac{I_0}{I}\right]_{\lambda 1} \quad (16)$$

$$\Delta A_{\lambda 2} = \Delta \left[\log \frac{I_0}{I}\right]_{\lambda 2} \quad (17)$$

Now, (15) can be expressed by combining (16) and (17); thus, the ratio can be calculated directly from the light received from the fingertip.

$$R = \frac{\Delta\left[\log \frac{I_0}{I}\right]_{\lambda 2}}{\Delta\left[\log \frac{I_0}{I}\right]_{\lambda 1}} = \frac{\left[\log \frac{I_0(d1)}{I(d1)} - \log \frac{I_0(d2)}{I(d2)}\right]_{\lambda 2}}{\left[\log \frac{I_0(d1)}{I(d1)} - \log \frac{I_0(d2)}{I(d2)}\right]_{\lambda 1}} = \frac{\left[\log \frac{I(d1)}{I(d2)}\right]_{\lambda 2}}{\left[\log \frac{I(d1)}{I(d2)}\right]_{\lambda 1}} \quad (18)$$

Table 1. Molar absorption coefficients.

Substance	Molar absorption coefficient ($cm^{-1} \cdot M^{-1}$)	
	$\lambda 1$ = 950 nm	$\lambda 2$ = 660 nm
Glucose (ε_G)	0.001	0.0002
HbO (ε_{HbO})	1204	319.6
HHb (ε_{HHb})	602.24	3226.56

The molar absorption coefficients of HbO, HHb, and glucose at two different wavelengths (660 and 950 nm) are listed in Table 1. The molar absorption coefficients of HbO and HHb were obtained from [17] and that of glucose was obtained from [18].

Using the molar absorption coefficient values, Equation (15) can be expressed as follows:

$$R = \frac{(0.0002 \times P_G) - (3226.56 \times P_G) - (3226.56 \times P_{HbO}) + 3226.56 + (319.6 \times P_{HbO})}{(0.001 \times P_G) - (602.24 \times P_G) - (602.24 \times P_{HbO}) + 602.24 + (1204 \times P_{HbO})} \quad (19)$$

The P_{HbO} can be obtained using (9) and (10) as

$$P_{HbO} = (1 - P_G) \times SpO_2 \quad (20)$$

By substituting (20) into (19),

In (21), there are two unknowns, namely P_G and SpO_2. R can be calculated using (18). By substituting the SpO_2 obtained from section 2.2 (or from reference), the value of P_G can be finally obtained. Therefore, the final equation of P_G can be expressed using (21) as follows:

$$R = \frac{(0.0002 \times P_G) - (3226.56 \times P_G) - (3226.56 \times (1 - P_G) \times SpO_2) + 3226.56 + (319.6 \times (1 - P_G) \times SpO_2)}{(0.001 \times P_G) - (602.24 \times P_G) - (602.24 \times (1 - P_G) \times SpO_2) + 602.24 + (1204 \times (1 - P_G) \times SpO_2)} \quad (21)$$

$$P_G = \frac{(R \times 602.24 \times SpO_2) - (R \times 602.24) - (R \times 1204 \times SpO_2) - (3226.56 \times SpO_2) + 3226.56 + (319.6 \times SpO_2)}{(R \times 0.001) + (R \times 602.24 \times SpO_2) - (R \times 602.24) - (R \times 1204 \times SpO_2) - (0.0002) - (3226.56 \times SpO_2) + 3226.56 + (319.6 \times SpO_2)} \quad (22)$$

From (7) and (8), we have

$$c_g = P_G \times c_T \tag{23}$$

where c_T = 150/64500 mol·dm^{-3} [12,19,20]. The value of c_T typically represents the molar concentration of whole blood and is used to calculate the blood glucose concentration.

After obtaining the desired values of P_G using both SPO_2 values from Section 2.2, and the reference SPO_2 separately, the blood-glucose concentration can be determined using (23). In both cases, the accuracy of the calculation of blood glucose concentration was compared and analyzed in the Results section.

2.2 SpO2 Calculation

To calculate SpO_2 values from the PPG signals, we followed the method described in [21]. The ratio R_{SpO_2} was calculated from the ratio of the normalized intensity of the received infrared light ($I_{n_{\lambda 1}}$) to red light ($I_{n_{\lambda 2}}$) and is expressed as (24).

$$R_{SpO_2} = \frac{\Delta A_{\lambda 2}}{\Delta A_{\lambda 1}} = \frac{ln\left(I_{n_{\lambda 2}}\right)}{ln\left(I_{n_{\lambda 1}}\right)} \tag{24}$$

As light passes through the additional optical path Δd at systole, from (11), it can be written as

$$d1 = d2 + \Delta d \tag{25}$$

The normalized intensity of the received light at a wavelength λ can be expressed as

$$I_{n_\lambda} = \frac{I}{I_{H_{d2}}} \tag{26}$$

where I represents the light intensity received by the photodetector (PD), and $I_{H_{d2}}$ represents the highest intensity during diastole.

The absorbance at wavelength λ can be found using the concentrations of oxyhemoglobin and deoxyhemoglobin as follows:

$$\Delta A_\lambda = (\varepsilon_{HbO}(\lambda) \times SpO_2 + \varepsilon_{HHb}(\lambda)(1 - SpO_2)) \times (c_{HbO} + c_{HHb}) \times \Delta d \tag{27}$$

Now, (23) can be expressed as

$$R_{SpO_2} = \frac{\Delta A_{\lambda 2}}{\Delta A_{\lambda 1}} = \frac{(\varepsilon_{HbO}(\lambda 2) \times SpO_2 + \varepsilon_{HHb}(\lambda 2)(1 - SpO_2)) \times (c_{HbO} + c_{HHb}) \times \Delta d}{(\varepsilon_{HbO}(\lambda 1) \times SpO_2 + \varepsilon_{HHb}(\lambda 1)(1 - SpO_2)) \times (c_{HbO} + c_{HHb}) \times \Delta d} \tag{28}$$

Finally, the oxygen saturation (SpO_2) can be calculated as

$$SpO_2 = \frac{\varepsilon_{HHb}(\lambda 2) - (\varepsilon_{HHb}(\lambda 1) \times R_{SpO_2})}{(\varepsilon_{HHb}(\lambda 2) - \varepsilon_{HbO}(\lambda 2)) + (\varepsilon_{HbO}(\lambda 1) - \varepsilon_{HHb}(\lambda 1) \times R_{SpO_2})} \tag{29}$$

The values depicted in Fig. 2(a) were used for SpO_2 to obtain the desired PG value of P_G from Equation (22).

2.3 Data Acquisition and Preprocessing

Our model was derived from the Beer-Lambert law, and states that the PPG signal should be transmissive. The finger is one of the best

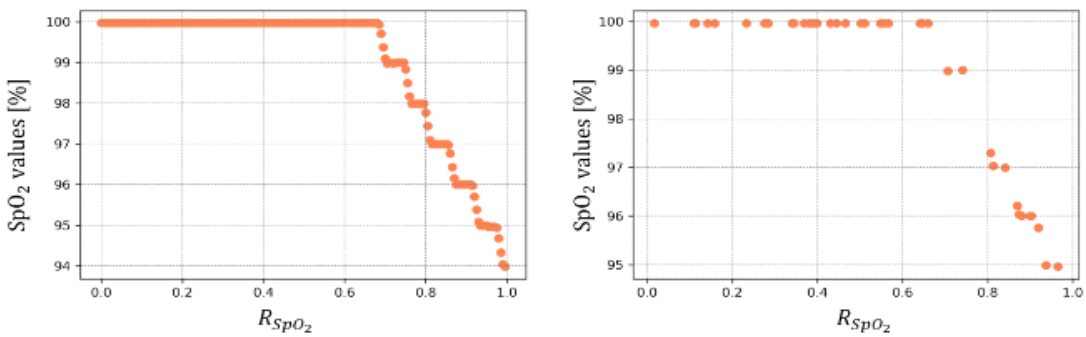


Fig. 2. SpO2 vs. ratio values (a) from [21] and (b) from our study (No. of subjects=40).

measurement sites for obtaining PPG waveforms[22]. Therefore, for transmissive PPG signal acquisition, the light source and sensor should be positioned on the opposite sides of the finger. In this case, high-intensity light is required to acquire high-quality PPG signals. To collect the transmissive-mode PPG signals for our model, a hardware system was developed using ESP32-PICO-V3 as the processing unit. This microcontroller was designed with an on-chip radio frequency (RF) communication system that eliminates the need for an external communication module to transmit data to a remote server. The SFH 7050 is a surface-mounted device (SMD) module containing three different LED wavelengths, namely green (525 nm), red (660 nm), and infrared (950 nm), along with a PD to collect transmissive-mode PPG signals from subjects. A biosensing analog front end (AFE), AFE 4404, was used to control the system. Fig. 3 shows a block diagram of the proposed hardware system.

To collect real PPG signals, 40 participants voluntarily agreed to provide their PPG signals. Written informed consent was obtained from the volunteers if they agreed to provide data for this study according to the Kookmin University IRB protocol. The volunteers were checked for severe disease within one month or irregular heart rate. They were excluded from the study if they had any of those two criteria. Of the 40 participants, 29 were male and 11 were female. Among the subjects, 22 were healthy and the remaining 18 were diabetic. Table 2 lists the statistical information of the participants' data. Statistical information includes the minimum (min), maximum (max), mean, and standard deviation (SD) of the data. For each subject, 240 s PPG signals were recorded at a sampling rate of 83 Hz. Because we developed our model for two LED wavelengths (950 and 660 nm), PPG signals recorded at infrared and red wavelengths were considered.

Table 2. Statistical information of the subjects' data.

Parameter	Age [years]	BMI
Min	25	20.8
Max	80	33.4
Mean	43.5	29.25
SD	18.6	2.76

We also measured the blood glucose concentrations and SpO2 values of the subjects using a CareSens II Plus system[23] and a Schiller Argus OXM Plus device [24] as a reference. The system accuracy results for the CareSens II Plus device were advertised as 99.7% within ± 15 mg/dL for blood glucose concentrations between 29.5 mg/dL and 455 mg/dL, and the stated accuracy for the Schiller Argus OXM Plus device was ± 2%, ranging from 70% to 100%. The subjects were instructed to remain stationary while recording the PPG signal to avoid external influences, such as movement and electromagnetic disturbances. Written consent was obtained from the participants for collection of blood glucose concentrations using an invasive device. Because we considered obtaining fasting blood glucose concentrations from the subjects, the subjects were asked not to eat for at least

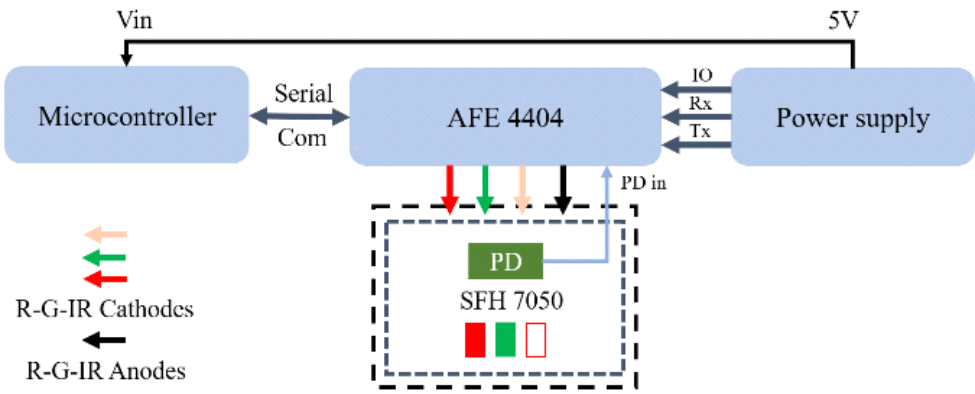


Fig. 3. Block diagram of the developed hardware system.

8 h before testing. We took the necessary precautions to collect blood glucose concentration data and strictly followed the instructions provided in the device manual to measure blood glucose concentration. Fig. 4 depicts an example of the PPG signals recorded from a subject, and Fig. 5 shows a histogram of the measured SpO2 and blood glucose concentrations of the subjects.

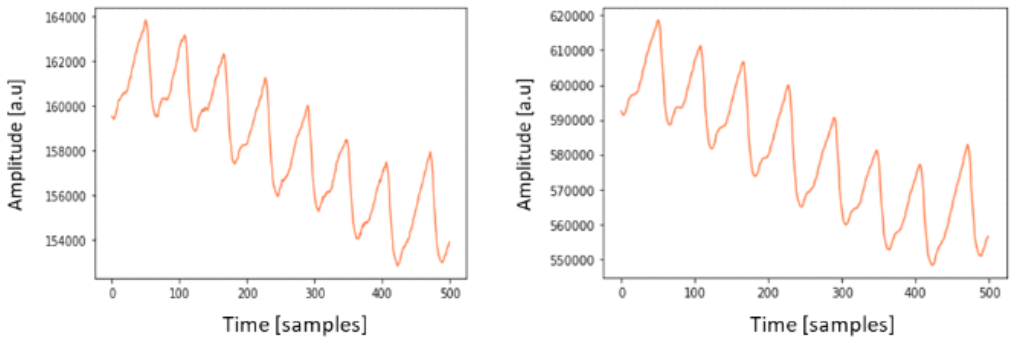
In the preprocessing stage, the PPG signals pass through several steps. To retain only the PPG signal, the high- and low-frequency noise, as well as the baseline drift, must be removed. The fitting-based sliding window (FSW) algorithm[6], which is convenient for detecting valleys, was implemented to remove baseline drift from PPG signals. A second-order low-pass Butterworth filter with a cutoff frequency of 8 Hz was used to remove high-frequency noise from the signals.

Using the filtered PPG signals, the ratio was calculated using Equation (18) for each participant. After calculating the PG value of P_G using equation (22), the blood glucose concentration (c_g) was calculated using equation (23) for each value of P_G. The units of the calculated blood-glucose concentrations (c_g) are in mmol·L^{-1}. To prepare the blood-glucose concentrations for calibration, they are converted to units of mg·dL^{-1} using (30).

$$1\ mg \cdot dL^{-1} = 0.0555\ mmol \cdot L^{-1} \tag{30}$$

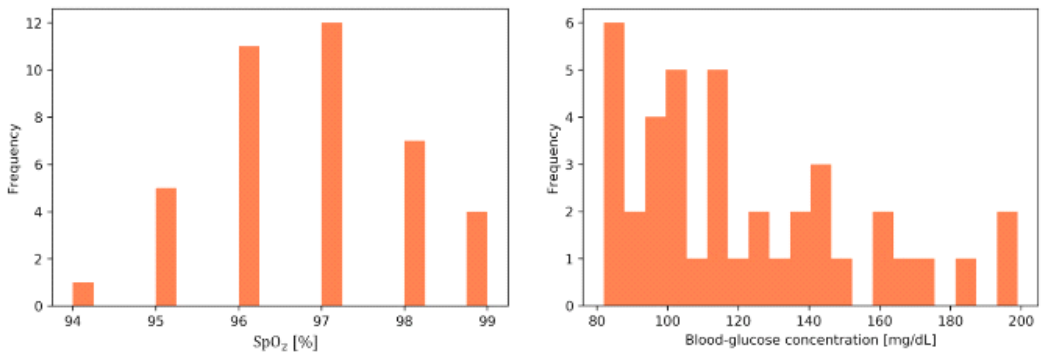
2.4 Calibration

A calibration step is required to derive more accurate estimates of blood glucose concentrations from (23). Calibration was performed on the calculated blood glucose concentrations by setting the measured concentrations as targets. A supervised machine learning model, XGBoost, with a learning rate of 0.3, maximum depth of six, and 100 estimators, was used for calibration. Leave-one-out cross validation (LOOCV) was performed to evaluate the



(a) (b)

Fig. 4. Examples of recorded PPG signals from a subject at (a) infrared and (b) red wavelengths.



(a) (b)

Fig. 5. Histograms of measured (a) SpO2 and (b) blood-glucose concentration values.

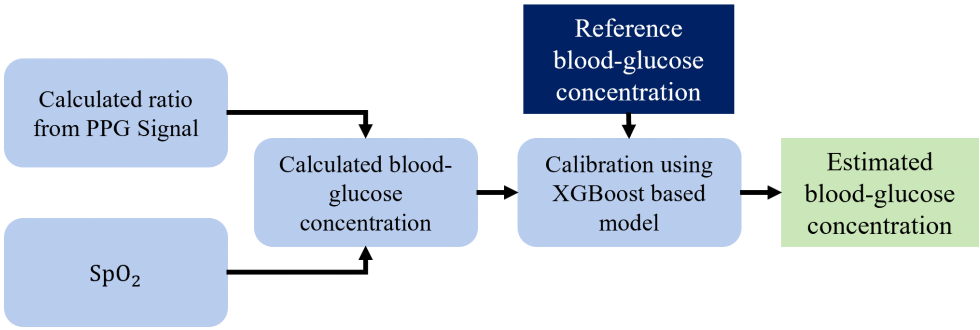


Fig. 6. Block diagram of the calibration process using XGBoost.

calibration results. LOOCV is a unique case of cross validation in which the number of folds depends on the number of instances in the training dataset. As our study focuses on estimating blood glucose concentrations from PPG signals using the Beer-Lambert-based model, implementing LOOCV in the regression will provide a reliable and unbiased estimate of model performance. While implementing LOOCV, 5% of the test data was used as the training set for the personalized regression model. This 5% of the data roughly corresponds to 114 PPG signal data samples per subject (approximately 1.4 seconds of PPG signal data). In [25], obtaining a personalized regression model using a percentage of the test data was considered an important step. We used the 1%, 2%, 3%, and 4% test data to determine the best results. Fig. 6 shows a block diagram of the calibration process using XGBoost.

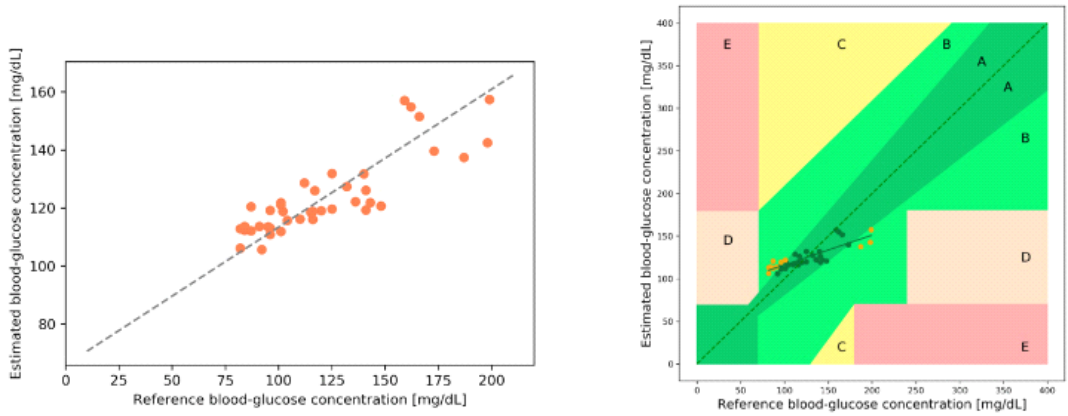
To evaluate the performance of the Beer-Lambert-law-based model in estimating blood glucose concentrations, three evaluation metrics were considered: the mean absolute error (MAE), root mean-squared error (RMSE), and Pearson correlation coefficient (Pearson's r). Clark error grid analysis (EGA)[26] and Bland–Altman plots were used to visualize the clinical accuracies of the estimated concentrations and estimation errors.

Ⅲ. Results

In this section, we present the evaluation of the proposed model compared with the reference data recorded from the volunteers for the considered evaluation metrics, EGA plots, and Bland–Altman analyses. In the proposed model, we estimated the blood glucose concentration using mathematical equations derived from the Beer-Lambert law, using the estimated SpO_2 values obtained from Section 2.2. To analyze the performance of the proposed model, we used the reference SpO_2 values to replace the estimated values when calculating the blood glucose concentrations. This analysis clearly depicts the comparable results produced by the proposed model. Finally, our model was compared with those in related studies to better illustrate its effectiveness. The following subsections present the results.

3.1 Estimation of Blood-Glucose Concentration using Estimated SpO2

The following results were obtained after calculating the blood glucose concentration using equations (21)-(23): To calculate the blood glucose concentration, we used the SpO_2 values calculated in Section 2.2. Fig. 7(a) depicts the fitted scatter plot after the calibration of the calculated blood glucose concentrations using XGBoost, and Fig. 7(b) illustrates the EGA plot of the estimated blood glucose concentrations compared with the reference blood glucose concentration values. The EGA is divided into five zones, A, B, C, D, and E, which are defined as follows: Zone A, clinically accurate data; Zone B, data outside 20% of the reference but cannot lead to inappropriate treatment; Zone C, data may lead to errors in treatment; Zone D, data lead to erroneous failures in treatment; and Zone E, data represent the potential for inappropriate treatment[26]. In our study, Zone A contained approximately 77.5% of the



(a) (b)

Fig. 7. (a) Fitted scatter plot after XGBoost regression; (b) error grid analysis (EGA) plot for the estimated blood-glucose concentration.

estimated blood glucose concentrations, and Zone B contained the remaining 22.5%. There were no samples in Zones C, D, or E. Table 3 shows the zonal accuracy of the EGA plot.

The Bland–Altman analysis shown in Fig. 8 indicates that the estimated blood-glucose concentration provides a bias of −2.14 ± 22.12 with limits of agreement (95%, 1.96 SD) ranging from -45.5 to +41.22. The values of the evaluation metrics are listed in Table 4.

Table 3. Zonal accuracies of the EGA plot.

Zone	A	B	C	D	E
Sample percentage	77.5%	22.5%	0%	0%	0%

Table 4. Evaluation metric values.

Metric	MAE	RMSE	Pearson's r
Values	16.26	21.22	0.85

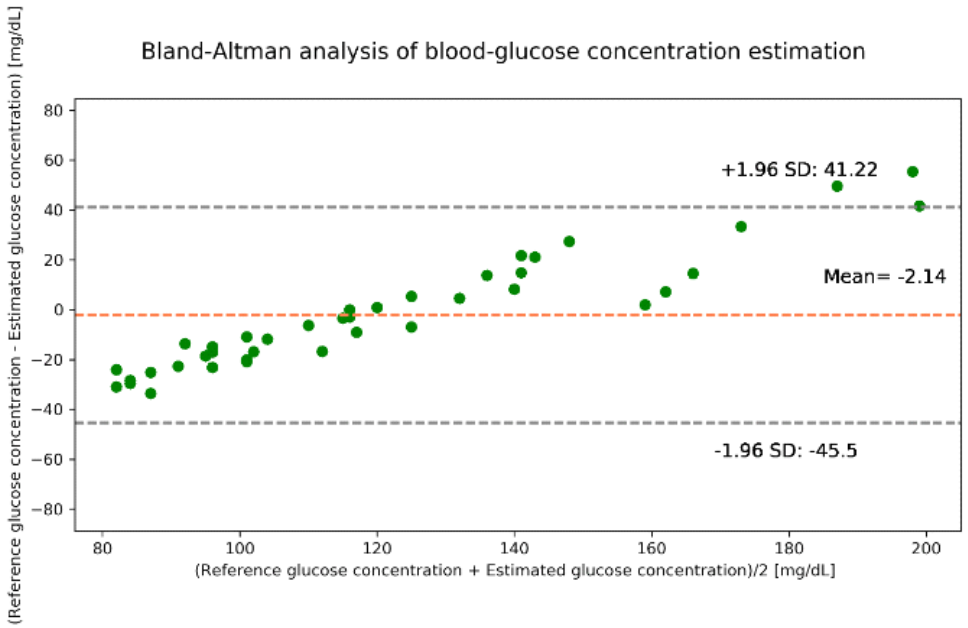


Fig. 8. Bland–Altman analysis for the estimated blood-glucose concentration using estimated SpO2.

3.2 Estimation of Blood-Glucose Concentration using Reference SpO2

The purpose of estimating blood glucose concentrations using the reference SpO_2 was to assess the performance variations of the proposed model. If the estimated blood glucose concentration using the reference SpO_2 is comparable (i.e., the performance variation is low) to that using the estimated SpO_2, then this proves the potential of our proposed model for application in real-world scenarios.

The following results were obtained after calculating the blood glucose concentrations using equations (21)-(23) with the reference SpO_2 values. Fig. 9(a) depicts the fitted scatter plot after the calibration of the calculated blood glucose concentration using XGBoost, and Fig. 9(b) illustrates the EGA plot of the estimated blood glucose concentration compared with the reference blood glucose concentration values. Zone A contains approximately 82.5% of the estimated blood glucose concentrations, and Zone B contains the remaining 17.5%. There were no samples in Zones C, D, or E. Table 5 shows the zonal accuracy of the EGA plot.

The Bland–Altman analysis shown in Fig. 10 indicates that the estimated blood-glucose concentrations provide a bias of -2.06 ±18.58 with the limits of agreement (95%, 1.96 SD) ranging from -38.47 to +34.36. The values of the evaluation metrics are listed in Table 6.

Table 5. Zonal accuracies of the EGA plot.

Zone	A	B	C	D	E
Sample percentage	82.5%	17.5%	0%	0%	0%

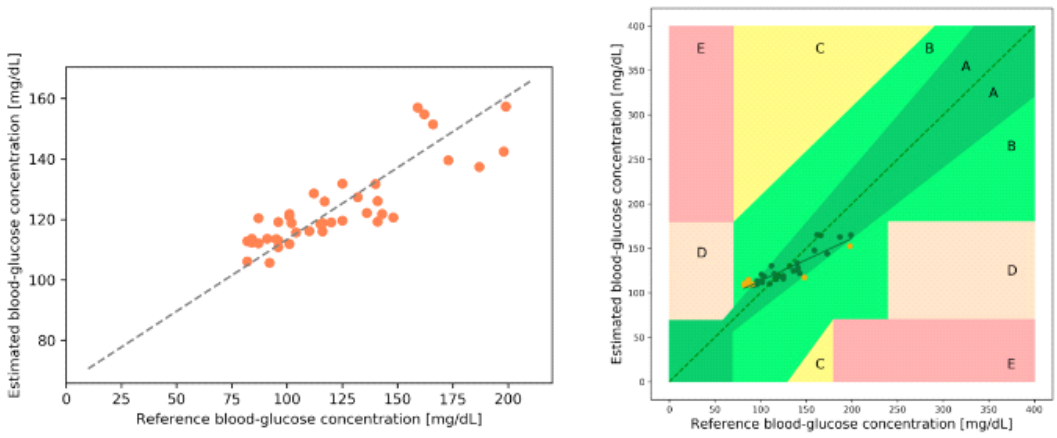
Table 6. Evaluation metric values.

Metric	MAE	RMSE	Pearson's r
Values	15.36	18.69	0.89

3.3 Comparisons with Other Related Works

Comparison analyses between the proposed model and other established models, as explained in [9] and [13], are presented in this subsection, considering identical scenarios for all models. We also present a performance analysis of the proposed model using reference SpO_2 values. Table 7 presents the evaluation metrics used for comparison.

In [9], the authors collected PPG signals in transmissive and reflective modes. As our model was evaluated with transmissive mode PPG signals, we considered the transmissive mode PPG signal for estimating the blood glucose concentration using the method described in [9]. After segmenting the individual PPG signals of 40 subjects for about 3 s for both red (660 nm) and infrared (950 nm) wavelengths, 17 discriminant features comprising a mixture of PPG-based physiological features, signal oriented characteristics, and physical parameters such



(a) (b)

Fig. 9. (a) Fitted scatter plot after XGBoost regression; (b) error grid analysis (EGA) plot for the estimated blood-glucose concentration.

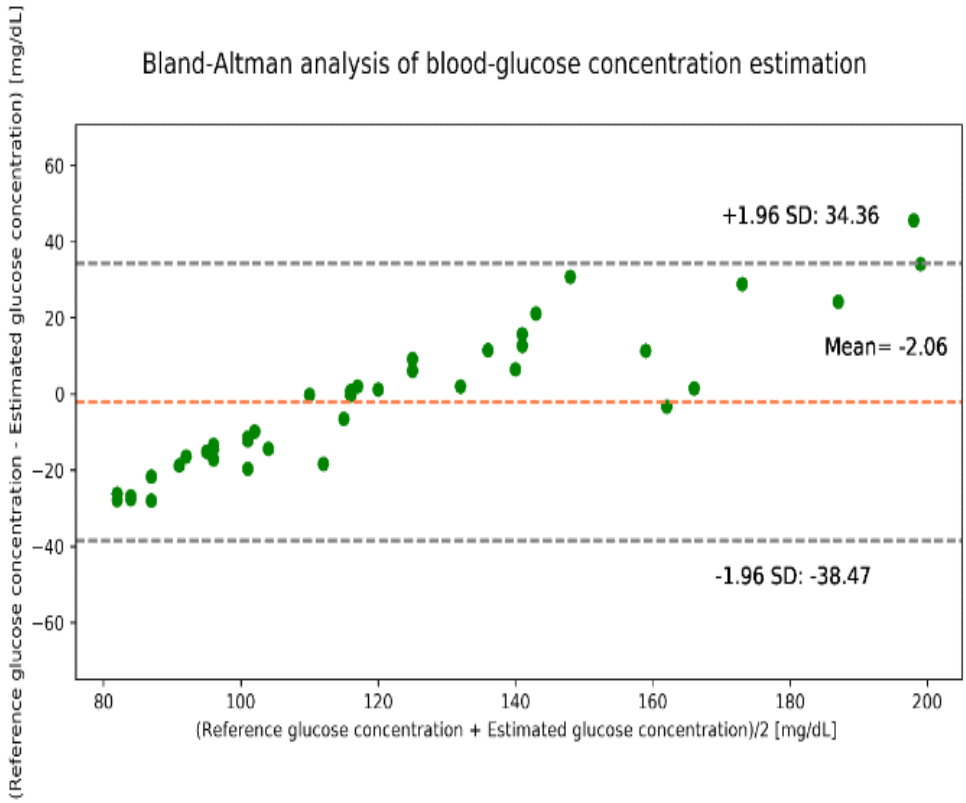


Fig. 10. Bland–Altman analysis for the estimated blood-glucose concentrations using reference SpO2.

as zero-crossing rate (ZCR), autocorrelation, Kaiser–Teager energy (KTE), power spectral density (PSD), autoregressive coefficients (ARC), blood oxygen saturation (SpO_2), and body mass index (BMI) were extracted for use as input features to the model while setting the measured blood-glucose concentrations as the targets. The feature vector equation for each frame f of the PPG signal s can be expressed as follows[9]:

In our recent study[13], we developed a finger model based on a Monte Carlo simulation of photon propagation.

From Table 7, it can be observed that the Pearson's r of our proposed model outperforms both the random forest and XGBoost models of [9] under the same scenario (i.e., with the estimated SpO_2). To estimate blood-glucose concentrations, the models in [9] require 17 features as input, whereas our proposed model requires no features, which reduces the complexity of our proposed model and makes it more efficient. We also used the estimated SpO_2 values to calculate blood glucose concentrations using the method described in our previous work[13], where the value of Pearson's r was found to be 0.84. After comparison with other noninvasive in vivo estimation methods of blood glucose concentrations, our proposed model showed relatively better accuracy in terms of Pearson's r value with less computational complexity, and the proposed method can be used for research purposes.

$$X_F^f = \left[\begin{array}{c} s_{zcr}, s_{ACR}, s_{PSD}^{kurt}, s_{PSD}^{var}, s_{PSD}^{mean}, s_{KTE}^{kurt}, s_{KTE}^{var}, s_{KTE}^{mean}, s_{KTE}^{skew}, s_{spec}^{kurt}, s_{spec}^{skew}, \\ s_{wavelet}^{mean}, s_{AR}, s_{spo2}, s_{skew}, s_{sad}, BMI \end{array} \right] \quad (31)$$

Table 7. Performance comparisons with the models in [9] and [13].

Model	Pearson's r	MAE	RMSE	Number of features	Calibration Step	Number of estimators
Random Forest [9]	0.71	16.4	20.1	17	0	1000
XGBoost [9]	0.77	11.4	15.92	17	0	100
MCS based model [13]	0.84	12.16	16.86	1	1	100
Proposed [with estimated SpO2]	0.85	16.26	21.22	0	1	0

Ⅳ. Discussion

A Beer-Lambert-law-based model for the noninvasive in vivo estimation of blood glucose concentration using PPG signals is presented in this paper. Considering blood as a homogeneous solution of hemoglobin and glucose, mathematical equations were derived from the Beer-Lambert law. The Beer-Lambert-law-based model was developed using two LED wavelengths: 950 (infrared) and 660 nm (red). The ratio was then calculated from the PPG signals for these two wavelengths, and a method for calculating the blood glucose concentration was developed. Transmission-mode PPG signals were collected from 40 subjects, along with their blood glucose concentrations and SpO_2 values as references.

While recording the PPG signals, the subjects were instructed to remain static and free from external influences such as movement and electromagnetic disturbances. Reference blood glucose concentrations and SpO_2 values were collected using the CareSens II Plus and Schiller Argus OXM Plus devices, respectively. From the PPG signals, the ratio values R and R_{SpO_2} were calculated and used to estimate blood glucose concentrations. A supervised machine learning model, XGBoost, was used to calibrate the estimated results, and Pearson's correlation coefficient (Pearson's r) was found to be 0.85. EGA plots and Bland–Altman analysis are presented to demonstrate the clinical accuracy of the estimations.

While calculating the blood glucose concentration using the reference SpO_2, Table 6 shows that Pearson's r is 0.89, whereas that for our proposed model with the estimated SpO_2 is 0.85. Although the performance of the proposed model with the reference SpO_2 is slightly better than that with the estimated SpO_2, the variation is very low and can be considered to have similar estimation results in real-world scenarios.

Previous studies on the noninvasive estimation of blood glucose concentrations[6,8-10] have been based on extracting time-domain, statistical, and signal-oriented features from PPG signals and using them to estimate blood glucose concentrations using machine learning algorithms. Along with the features of PPG signals, physical parameters, such as BMI, height, weight, and age, have also been used in these studies, which often create obstacles for estimating blood glucose concentrations more accurately. To reduce these complexities from the features extracted from the PPG signals and the possibility of inaccurate estimation, we derived mathematical equations based on the Beer-Lambert law to estimate the blood glucose concentration.

Our proposed model is derived by considering a simple blood vessel model that neglects some effects such as pressure at the PPG signal measurement site, skin type, and finger width. However, if these effects are compensated when designing the model, higher accuracies may be obtained for the estimated blood glucose concentrations. Furthermore, our model was tested using a small dataset with fewer variations between subjects. Therefore, future studies should consider the effects of these variabilities using more diverse datasets, including both hypoglycemic and hyperglycemic subjects. Despite the limitations of our proposed model, it performs better than other compared methods, and at this stage, it can be said to be suitable for self-monitoring applications of a person.Figure 1

Ⅴ. Conclusion

In this study, we proposed a Beer-Lambert law-based model for estimating blood glucose concentrations. In the proposed model, the ratio was calculated from the PPG signals of the subjects to obtain the blood glucose concentration using mathematical equations. XGBoost, a supervised machine learning model, was used to calibrate the estimated blood glucose concentrations with reference concentrations. The results were evaluated using Pearson's r, error grid analysis (EGA) plots, and Bland-Altman analysis. Compared with other notable noninvasive blood glucose estimation models, our proposed model was less complex and had a higher accuracy. The device developed in the current study can be used as a point-of-care monitoring device on a daily basis, which can alleviate the patient's burden of checking blood glucose concentration multiple times using an invasive device.

Funding: This work was supported by the National Research Foundation (NRF) of Korea funded by the Ministry of Science and ICT (2022R1A5A7000765) and by the Basic Science Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Education (NRF-2022R1A2C2010298). This research was supported by the Korea Industrial Technology Association (KOITA), funded by the Ministry of Science and ICT (MSIT).

Author's Contributions: Chowdhury Azimul Haque: Conceptualization, Methodology, Software, Writing-Original draft; **Shifat Hossain:** Methodology, Software, Validation; **Tae-Ho Kwon**: Formal analysis; **Hyoungkeun Kim:** Resources, Validation **Ki-Doo Kim:** Validation, Supervision, Project administration, and funding acquisition.

Data Availability Statement: We dataset was created for this study. As further research is in progress, we cannot publish the dataset immediately.

Ethical Statement: All protocols and procedures in this study were approved by the Institutional Review Board (IRB) of Kookmin University, Seoul, Korea (approval date:17th July 2020). The procedures followed the Helsinki Declaration of 1975 and were revised in 2008. All participants agreed to participate and share their data for academic research purposes. The IRB protocol number was KMU-202006-HR-237.

Conflicts of Interest: The authors declare that they have no conflicts of interest. The funders had no role in the design of the study; collection, analyses, or interpretation of data; writing of the manuscript; or publication of results.

References

[1] American Diabetes Association, "*Diabetes Symptoms, Causes, & Treatment*," 2021, Accessed Sep. 27, 2021, https://www.diabetes.org/diabetes.

[2] L. Heinemann, "Finger pricking and pain: A never ending story," *J. Diabetes Sci. Technol.*, vol. 2, no. 6, pp. 919-921, 2008. (https://doi.org/10.1177/193229680800200526)

[3] X. Xing, et al., "An unobtrusive and calibration-free blood pressure estimation method using photoplethysmography and biometrics," *Sci. Rep.*, vol. 9, 8611, 2019. (https://doi.org/10.1038/s41598-019-45175-2)

[4] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. R1-R39, 2007. (https://doi.org/10.1088/0967-3334/28/3/R01)

[5] E. Monte-Moreno, "Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques," *Artif. Intell. Med.*, vol. 53, pp. 127-138, 2011. (https://doi.org/10.1016/j.artmed.2011.05.001)

[6] G. Zhang, et al., "A noninvasive blood glucose monitoring system based on smartphone PPG signal processing and machine learning," *IEEE Trans. Ind. Inform.*, vol. 16, pp. 7209-7218, 2020. (https://doi.org/10.1109/TII.2020.2975222)

[7] S. Hossain, et al., "Estimation of blood glucose from PPG signal using convolutional neural network," in *Proc. 2019 IEEE Int. Conf. BECITHCON*, pp. 53-58, 2019.

[8] C. Tsai, et al., "Diabetes care in motion: Blood glucose estimation using wearable devices," *IEEE Consum. Electron. Mag.*, vol. 9, pp. 30-34, 2020. (https://doi.org/10.1109/MCE.2019.2941461)

[9] S. Sen Gupta, et al., "Towards non-invasive blood glucose measurement using machine learning: An all-purpose PPG system design," *Biomed. Signal Process. Control*, vol. 68, 102706, 2021. (https://doi.org/10.1016/j.bspc.2021.102706)

[10] V. P. Rachim and W.-Y. Chung, "Wearable-band type visible-near infrared optical biosensor for non-invasive blood glucose monitoring," *Sens. Actuators B Chem.*, vol. 286, pp. 173-180, 2019. (https://doi.org/10.1016/j.snb.2019.01.121)

[11] P. P. Banik, et al., "Development of a wearable reflection-type pulse oximeter system to acquire clean PPG signals and measure pulse rate and SpO2 with and without finger motion," *Electronics*, vol. 9, no. 11, 1905, 2020. (https://doi.org/10.3390/electronics9111905)

[12] S. Hossain, et al., "Derivation and validation of gray-box models to estimate noninvasive in-vivo percentage glycated hemoglobin using digital volume pulse waveform," *Sci. Rep.*, vol. 11, 12169, 2021. (https://doi.org/10.1038/s41598-021-91527-2)

[13] C. A. Haque, et al., "Noninvasive in vivo estimation of blood-glucose concentration by monte carlo simulation," *Sensors*, vol. 21, no. 14, 4918, 2021. (https://doi.org/10.3390/s21144918)

[14] T. Tamura, et al., "Wearable photoplethysmo-graphic sensors — past and present," *Electronics*, vol. 3, no. 2, pp. 282-302, 2014. (https://doi.org/10.3390/electronics3020282)

[15] J. Yadav, et al., "Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy," *Biomed. Signal Process. Control*, vol. 18, pp. 214-227, 2015. (https://doi.org/10.1016/j.bspc.2015.01.005)

[16] *The Trouble with Properly Describing the Oxygen-Transport-Related Quantities*, Available online: https://acutecaretesting.org/en/articles/the-trouble-with-properly-describing-the-oxygentransportrelated-quantities (accessed on 13 Aug. 2021).

[17] *Tabulated Molar Extinction Coefficient for Hemoglobin in Water*, Available online: https://omlc.org/spectra/hemoglobin/summary.html?fbclid=IwAR2F5VOmtzakCum3DpZYBps3ZC3qsAkSjnCVsRVKmac-gV6D4ha4lJY_1Iw (accessed on 11 Jan. 2021).

[18] Andor Glucose Spectrum | Measuring the Spectral Response of Glucose.

[19] *Anonymous Diagnosis*, Available online: https://www.llscanada.org/myeloproliferative-neoplasms/polycythemia-vera/diagnosis (accessed on 1 Oct. 2021).

[20] M. C. P. Van Beekvelt, et al., "Performance of near-infrared spectroscopy in measuring local O2 consumption and blood flow in skeletal muscle," *J. Appl. Physiol.*, vol. 90, no. 2, pp. 511-519, 2001. (https://doi.org/10.1152/jappl.2001.90.2.511)

[21] C. M. Lochner, et al., "All-organic optoelectronic sensor for pulse oximetry," *Nat. Commun.*, vol. 5, 5745, 2014. (https://doi.org/10.1038/ncomms6745)

[22] V. Hartmann, H. Liu, F. Chen, Q. Qiu, S. Hughes, and D. Zheng, "Quantitative comparison of photoplethysmographic waveform characteristics: Effect of measurement site," *Front. Physiol.*, vol. 10, 198, 2019. (https://doi.org/10.3389/fphys.2019.00198)

[23] i-SENS, "*CareSens® II Plus Manual* (English)," 2019.

[24] www.Schiller.Ch, Available online: https://www.schiller.ch/bn/en/product/argus-oxm-plus (accessed on 21 Jun. 2021).

[25] G. Slapničar, N. Mlakar, and M. Luštrek, "Blood pressure estimation from

photoplethysmogram using a spectro-temporal deep neural network," *Sensors*, vol. 19, no. 15, 3420, 2019. (https://doi.org/10.3390/s19153420)

[26] W. L. Clarke, "The original clarke error grid analysis (EGA)," *Diabetes Technol. Ther.*, vol. 7, no. 5, pp. 776-779, 2005. (https://doi.org/10.1089/dia.2005.7.776)

초두리 아지물 하크 (Chowdhury Azimul Haque)

2019년 : Khulna University of Engineering and Technology (KUET) 전자공학과 학사
2022년 : 국민대학교전자공학과 석사
<관심분야> 디지털통신, 디지털 신호처리
[ORCID:0000-0002-3947-1454]

시팟 호세인 (Shifat Hossain)

2017년 : Khulna University o fEngineering and Technology (KUET) 전자공학과 학사
2021년 : 국민대학교 전자공학과 석사
2022년 : 국민대학교 전자공학과 연구원
2022년~현재 : Department of Electrical and Computer Engineering, University of Central Florida, Ph.D candidate
<관심분야> 디지털통신, 디지털신호처리
[ORCID:0000-0002-4537-2620]

권 태 호 (Tae-Ho Kwon)

2015년 : 국민대학교 전자공학과 학사
2018년 : 국민대학교 전자공학과 석사
2022년 : 국민대학교 전자공학과 박사
2022년~현재 : 국민대학교 멀티미디어 신호처리 연구실 책임연구원
<관심분야> 디지털통신, 디지털신호처리
[ORCID:0000-0001-6784-5591]

김 형 근 (Hyoungkeun Kim)

1995년 : 국민대학교 전자공학과 학사
1997년 : 국민대학교 전자공학과 석사
1998~1999년 : ㈜삼성전자 주임연구원
1999~2006년 : ㈜네이스텍 대표
2016년~2019년 : ㈜인디프로그 대표
2019년~현재 : ㈜한국아이티에스 연구소장
<관심분야> 디지털통신, 디지털신호처리

김 기 두 (Ki-Doo Kim)

1980년 : 서강대학교 전자공학과 학사
1980년~1985년 : 국방과학연구소 연구원
1988년 : 펜실베니아 주립대학교 전자공학 석사
1990년 : 펜실베니아 주립대학교 전자공학 박사
1998년~1999년 : 미국 UCSD, Visiting Scholar
1991년~2023년 2월 : 국민대학교 전자공학부 교수
2023년 3월~현재 : 국민대학교 전임연구교수
<관심분야> 디지털통신, 디지털신호처리
[ORCID:0000-0001-5052-3844]