

## Article

# Camera-Based Vital Sign Estimation Techniques and Mobile App Development

Tae Wuk Bae <sup>1,\*</sup>, Young Choon Kim <sup>2,\*</sup>, In Ho Sohng <sup>2</sup> and Kee Koo Kwon <sup>1</sup>

<sup>1</sup> Daegu-Gyeongbuk Research Center, Electronics and Telecommunications Research Institute, Daegu 42994, Republic of Korea; kwonkk@etri.re.kr

<sup>2</sup> Department of Information and Communication Security, U1 University, Asan 31415, Republic of Korea; songinho@u1.ac.kr

\* Correspondence: twbae@etri.re.kr (T.W.B.); yckim@yd.ac.kr (Y.C.K.); Tel.: +82-53-670-8063 (T.W.B.)

## Abstract

In this paper, we propose noncontact heart rate (HR), oxygen saturation ( $\text{SpO}_2$ ), and respiratory rate (RR) detection methods using a smartphone camera. HR frequency is detected through filtering after obtaining a remote PPG (rPPG) signal and its power spectral density (PSD) is detected using color difference signal amplification and the plane-orthogonal-to-the-skin method. Additionally, the  $\text{SpO}_2$  is detected using the HR frequency and the absorption ratio of the G and B color channels based on oxyhemoglobin absorption and reflectance theory. After this, the respiratory frequency is detected using the PSD of rPPG through respiratory frequency band filtering. For the image sequences recorded under various imaging conditions, the proposed method demonstrated superior HR detection accuracy compared to existing methods. The confidence intervals for HR and  $\text{SpO}_2$  detection were analyzed using Bland–Altman plots. Furthermore, the proposed RR detection method was also verified to be reliable.

**Keywords:** camera; digital imaging; heart rate; respiratory rate; oxygen saturation; remote PPG



Academic Editors: Elias N. Zois and Dimitrios Kalivas

Received: 10 June 2025

Revised: 27 July 2025

Accepted: 29 July 2025

Published: 31 July 2025

**Citation:** Bae, T.W.; Kim, Y.C.; Sohng, I.H.; Kwon, K.K.

Camera-Based Vital Sign Estimation Techniques and Mobile App Development. *Appl. Sci.* **2025**, *15*, 8509. <https://doi.org/10.3390/app15158509>

**Copyright:** © 2025 by the authors.

Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Recently, research has been conducted using camera images to extract the change components caused by the expansion and contraction of blood vessels as noncontact remote PPG (rPPG) signals. In a reflection model of an rPPG signal [1], some of the ambient light incident on exposed skin, such as that of the face, is specularly reflected from the skin's surface, and the remaining light is absorbed by the skin (see Supplementary Figure S1). Some of the absorbed light is reflected by blood vessels and exits the skin through diffuse reflection. The regular component of reflection from the skin's surface and diffuse reflection component, including the pulsation of the blood flow in the blood vessels, are acquired as camera images. The rPPG signal is extracted from the obtained image, and the heart rate (HR) is extracted from the diffuse reflection component caused by the pulsation of the blood vessels.

The methods for obtaining the HR using an RGB camera can be classified into blind source separation (BSS)- and model-based methods. BSS-based methods can be further classified into independent component analysis (ICA) [2,3] and empirical mode decomposition [4]. Poh et al. [2] and Kwon et al. [3] detected an area of a face as a region of interest in RGB images and extracted the average signal of each RGB channel in the frame units. The obtained RGB signal was converted into three separate, uncorrelated sources

through ICA, and a second signal with a separate source was used as an HR waveform. Chen et al. [4] calculated the HR by counting the peaks of the fourth intrinsic mode function (IMF) obtained using ensemble empirical mode decomposition (EEMD) from the G-channel reflection in a forehead ROI. The ICA method identifies the HR by separating the RGB signals from an image of a face into independent, uncorrelated signals, whereas the model-based method [5] uses a color model for extracting rPPG signals from color signals. In particular, the model-based method uses the minute skin color changes in the face due to the blood flow generated by a heartbeat. Most model-based methods use color models, which employ G [5], green–red difference (G–R or GRD) [5], adaptive GRD (AGRD) [6], LAB [7], skin color stratification [8], chrominance (CHROM) [9], and plane-orthogonal-to-the-skin (POS) methods [10]. The G method estimates the HR using only the G-channel signal, which is affected the most by the hemoglobin absorption in the blood [5]. In the G–R or GRD methods, the rPPG signal is obtained by subtracting the R-channel signal, which changes the least, from the G-channel signal, which changes the most significantly. In addition, the AGRD method uses the adaptive color difference between the green and red channels and an adaptive bandpass filter (BPF) to remove residual motion artifacts from the rPPG signal [6]. In the LAB method [7], EEMD and an IMF are applied to the difference between the A and B channels in the LAB color space, while the skin color amplification method [8] detects the HR by amplifying the skin color in an RGB image of the ROI. The CHROM method [9] extracts rPPG signals using a combination of color difference signals from the RGB channels. The POS method [10] creates new projection axis coordinates by combining three orthogonal R, G, and B coordinates and projects temporally normalized R, G, and B values onto the new projection axis to create a new two-dimensional color difference signal.

Recently, Chen et al. [11] proposed a method for estimating heart rate from rPPG signals under irregular frame rates and data loss conditions using low-complexity timing correction techniques (LCTC), including linear, cubic, and filter interpolation. In addition, Sun et al. [12] introduced a 3D Convolutional Neural Network (3D CNN) designed to consider both long-term periodic characteristics and short-term temporal features for effective rPPG signal extraction. However, although these recent methods demonstrate good performance, their effectiveness on video sequences captured under various lighting conditions has not been evaluated.

Because the conventional contact-based pulse oximeter measurement method uses a clip-on or fixed sensor device, it causes discomfort to patients. In addition, because COVID-19 patients require frequent and accurate blood oxygen saturation ( $\text{SpO}_2$ ) monitoring, the demand for noncontact measurement is increasing. Recently, a noncontact  $\text{SpO}_2$  measurement method was proposed, based on rPPG signals extracted from images of human faces [13–15].

Changes in the blood concentration due to the amount of oxygen transported by hemoglobin in red blood cells result in changes in the wavelength of the reflected light captured by a photodiode or camera. This is because the absorption characteristics of the light wavelengths of oxyhemoglobin ( $\text{HbO}_2$ ) and deoxyhemoglobin (Hb) are different. At each illumination wavelength, the absorption of Hb differs from that of  $\text{HbO}_2$ , and the  $\text{SpO}_2$  is estimated by measuring the absorption of blood at both wavelengths. Among the various light wavelengths, red (R, 660 nm) and IR (940 nm) light penetrate tissue and blood well; therefore, red and IR light are primarily used in PPG sensors. Alharbi et al. [13] measured the  $\text{SpO}_2$  using an optical filter that only let through green and orange wavelengths in front of a camera. Al-Naji et al. [14] measured the  $\text{SpO}_2$  from R and G images obtained using a general RGB camera without a separate optical filter. And Akamatsu et al. [15] proposed

a CNN-based SpO<sub>2</sub> detection technique that utilizes DC and AC components from RGB signals of facial images.

Recently, since the methods of simultaneously measuring the HR and respiratory rate (RR) using only an RGB or thermal camera are unreliable, bio-signal detection studies using both visible and thermal cameras have been conducted. Negish et al. [16] proposed a noncontact bio-signal measurement method in which the HR was extracted from a green (G) image of the face ROI taken using an RGB camera and the RR was extracted from the thermal signal of the area around the nose obtained using an infrared thermal camera. However, this method is not suitable for long-term respiration monitoring because the nose region is not always in the thermal camera's field of view.

Zhao et al. [17] proposed a method of measuring the HR from images of the face area and the RR from images of the chest area taken using a single RGB camera. Massaroni et al. [18] measured the RR by analyzing the change in brightness over time in the chest area using an RGB camera. In the chest region, intensity signals were extracted by averaging all the pixels in the vertical direction for the RGB channels. Since these signals included motion components caused by respiration and noise, a motion vector was obtained using an optical flow method to remove noise and extract the motion components. The time-varying motion vector was converted into the power spectral density (PSD), and after using a BPF for the 0.05 to 2.0 Hz frequency band corresponding to respiration, the frequency with the maximum amplitude was measured as the RR.

Table 1 shows a comparison of the experimental conditions used when testing various noncontact HR detection methods. Except for in tests of the proposed method, only one device was used, and all experiments except those using the POS and proposed methods were conducted for only one lighting situation and a limited age group. Our study used images taken using various mobile devices and under various lighting conditions to ensure the universality of the research results. Therefore, our approach lays important groundwork for ensuring that future noncontact healthcare algorithms maintain consistent accuracy when using various devices and under various lighting conditions.

**Table 1.** Comparison of experimental conditions used when testing various noncontact HR detection methods.

Methods	Image Sequence Used	Devices Used	Illumination
ICA [2,3]	640 × 480 pixels, 24 bit 20 s (15 frames/s) 12 subjects (aged from 18 to 31)	 MacBook Pro	Indoors, sunlight
AGRD [6]	640 × 480 pixels, 24 bit 20 s (30 frames/s) 10 subjects (aged from 20 to 33)	 Logitech C270	Fluorescent lamp
CHROM [9]	1024 × 752 pixels, 8 bit 25 s (20 frames/s) 117 subjects (unknown ages)	 USB UI-2230SE-C by IDS	(Studio illumination)
POS [10]	768 × 576 pixels, 8 bit 1125~2700 s (20 frames/s) 5~16 subjects according to experiments 60 video sequences (unknown ages)	 USB UI-2230SE-C by IDS	Fluorescent lamp Red, green, blue, red-green, red-blue, and green-blue LED lamps
Proposed	640 × 480, 852 × 480, 1280 × 720, 1080 × 1920 7 s (30 frames/s) 20 subjects, 508 video sequences (aged from 31 to 61)	 Galaxy A50, A9, Note 9, and S21 Plus, iPhone X, Logitech C270, MS HD-3000, LG Gram, MacBook Air	3200 K and 5600 K fluorescent lamps, sunlight

This paper proposes a non-contact method for detecting HR, SpO<sub>2</sub>, and RR by combining the POS method with a color difference signal amplification technique. Additionally, a time-shifted moving average process is proposed to eliminate temporal noise such as lighting flicker, along with a remeasurement algorithm based on heart rate comparison over time intervals to improve HR detection accuracy. The proposed method was validated for robustness against variations in image quality and lighting using video sequences recorded with various types of smartphone cameras and laptop webcams.

This paper is organized as follows. Section 2 introduces the contactless detection methods for HR, SpO<sub>2</sub>, and RR based on smartphone camera image sequences. Section 3 describes the recorded image sequence database and verifies the detection performance of the proposed method using Bland–Altman plots. It also presents the developed mobile application and analyzes the time cost and noise effects associated with the proposed method. Section 4 introduces the challenges related to cameras and video in non-contact bio-signal research, the lack of diverse ethnic databases, and future plans.

## 2. Materials and Methods

This section describes the experimental procedures for performing noncontact HR and SpO<sub>2</sub> measurements using mobile device cameras. The novel methods proposed were verified through a comparison with reference signals from contact devices. The present study's protocol was reviewed and approved by the Institutional Review Board of the Electronics and Telecommunications Research Institute (approval code N01-202304-01-006). Informed consent was provided by all the subjects when they were enrolled.

### 2.1. HR Detection

We built a noncontact HR measurement system using mobile device cameras (see Supplementary Figure S2). Various mobile devices were used in the experiment. To verify the HR estimated by the camera, HRs were measured simultaneously using a fingertip pulse oximeter, CHOCEMMED's MD300C22, Beijing, China.

The images used in the experiment were captured using various smart devices and under various lighting conditions as shown in Table 2. The variables included the illumination environment, acquisition device, image size, and number of frames. Various types of smartphone cameras, laptop cams, and webcams were used as image acquisition devices, and 20 subjects aged 21 to 61 participated in the experiment. The participants stood about 60–90 cm away from the camera and stared at it for about 30 s, with the video frame rate set to 30 fps. To measure the respiratory rate, guided breathing and spontaneous breathing methods were used alternately. Guided breathing meant that the participant breathed according to the breathing pattern displayed on a PC monitor, and spontaneous breathing meant breathing voluntarily regardless of the breathing pattern.

The proposed process for the noncontact measurement of the HR using a smartphone camera is shown in Figure 1. The face region was detected in real time using the AdaBoost algorithm [19] and set as the ROI. The proposed HR detection method combines color difference signal amplification and the POS method and is called the CA-POS method in this paper. The RGB image of the facial ROI was converted into color difference signals,  $Cb$  and  $Cr$ , which were then amplified [8]. The amplified  $Cb'$  and  $Cr'$  were calculated as follows.

$$Cb' = Cb_c + S \times (Cb - Cb_c) \quad (1)$$

$$Cr' = Cr_c + S \times (Cr - Cr_c) \quad (2)$$

where  $Cb_c$  and  $Cr_c$  represent the average color difference signals of the facial region, and  $S$  represents a scale.

**Table 2.** Characteristics of the acquired image sequences used in experiments.

Features	Images	Features	Images	Features	Images
Sunlight Galaxy A50 1080 × 1920 96 frames		Sunlight Galaxy Note 9 1080 × 1920 565 frames		5600 K LG Gram cam 1280 × 720 467 frames	
Sunlight iPhone X 1080 × 1920 381 Frames		Sunlight Macbook Air cam 640 × 480 780 frames		Sunlight LG Gram cam 1280 × 720 469 frames	
Fluorescent MS Webcam HD-3000 640 × 480 673 frames		Sunlight Logitech Webcam C270 1280 × 720 647 frames		3200 K Galaxy A9 852 × 480 404 frames	
Sunlight 1080 × 1920 S note 9 465 frames		LED Logitech Webcam C270 1280 × 720 606 frames		5600 K Galaxy A9 852 × 480 411 frames	
Fluorescent Galaxy S21 Plus 1080 × 1920 524 frames		3200 K LG Gram cam 1280 × 720 486 frames		Sunlight Galaxy A9 852 × 480 376 frames	

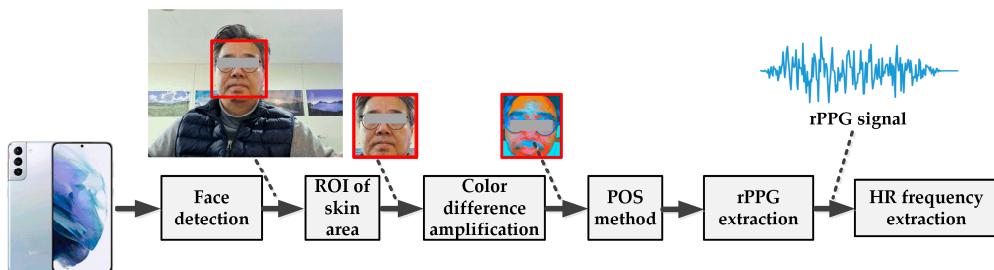
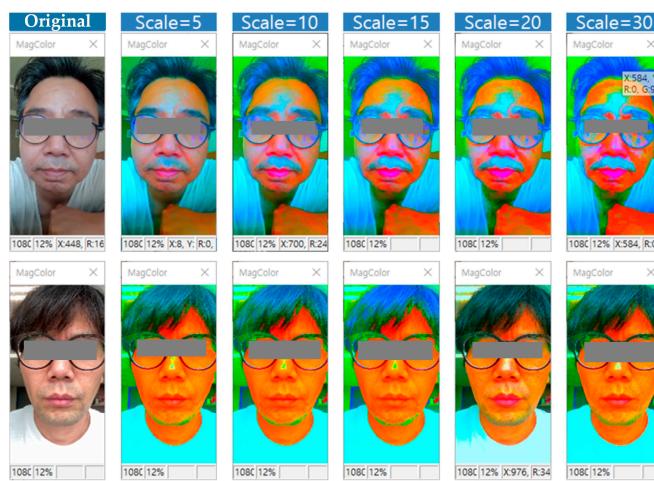
**Figure 1.** Proposed smartphone camera-based HR measurement process.

Figure 2 shows the changes in color of the images according to the amplification of the color difference signals. It can be seen that the color component of the image increased as the amplification increased. In general, when noise remains constant, increasing the scaling of the color difference signal leads to an improvement in the signal-to-noise ratio (SNR). However, if both the color difference signal scaling and the noise increase simultaneously, the SNR remains nearly unchanged. Close-up images captured with modern smartphone cameras typically yield high SNRs when there is minimal motion. Nevertheless, excessive scaling of the color difference signal may result in color saturation. In this study, the optimal scaling value ( $S = 10$ ) was empirically determined within a range that avoids color saturation. Then, the average RGB time signal per frame was calculated by averaging the signal from each RGB channel for the detected face regions. The rPPG signals were detected using the extracted RGB signals per frame and algorithms for the CHROM and POS techniques, which are representative HR estimation methods. Subsequently, the heartbeat frequency was extracted through a frequency analysis of the rPPG signal.



**Figure 2.** Changes in color of images according to amplification of color difference signals.

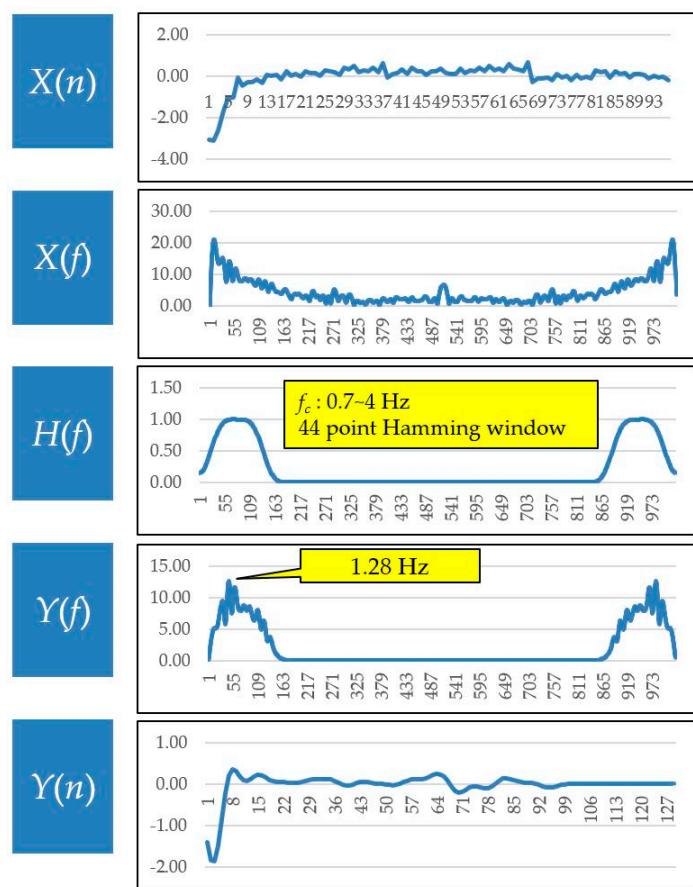
A real-time face detection experiment was performed using a smartphone camera, and images of various sizes including the face were acquired. The HR detection results were confirmed for 508 image sequences acquired over approximately 7 s (210 frames), and the images featured various backgrounds, lighting conditions, distances from the camera, and genders. To verify the accuracy of the results, the noncontact HR was compared with the HR measured using a finger blood volume pulse sensor or smart watch.

The mean value is calculated from the sum of the amplified chrominance signals  $a$  and  $b$  across all pixels within the facial region of interest (ROI). The filter sequence for HR extraction is shown in Figure 3. Zero padding and the FFT were performed on the signal  $x(n)$ , in which the average values of all the pixels in the ROI were arranged on a time axis. Subsequently, a BPF,  $H(f)$ , with a passband of 0.7 to 4.0 Hz and a Hamming window was used to detect the HR frequency (see Supplementary Figure S3a,b). The Hamming window maintains frequency resolution while reducing spectral leakage to some extent, due to its moderate main lobe width and effective side lobe suppression. Therefore, PSD analysis is widely used, in general, for balanced performance. The frequency with the peak value in the spectrum resulting from BPF processing was determined to be the HR frequency. For example, in one case the detected HR frequency was 1.28 Hz.

To improve the accuracy of HR detection, we divided the input signal by the time interval; when the difference between each detected HR was significant, the measurement was repeated (see Supplementary Figure S4). An example of the decision whether to accept an HR value or repeat the measurement is presented in Table 3. If the HRs measured for each time interval were similar, then they were considered accurate HRs; on the other hand, if the difference between the measured HRs was significant, then the measurement was repeated.

**Table 3.** Example of decision whether to accept HR value or repeat measurement.

		$X_1(n)$	$X_2(n)$	$X_3(n)$	$X_4(n)$
Ex (1)	Detected HR	96	97	96	96
	Final HR			96	
Ex (2)	Detected HR	96	107	97	96
	Final HR			Remeasurement	

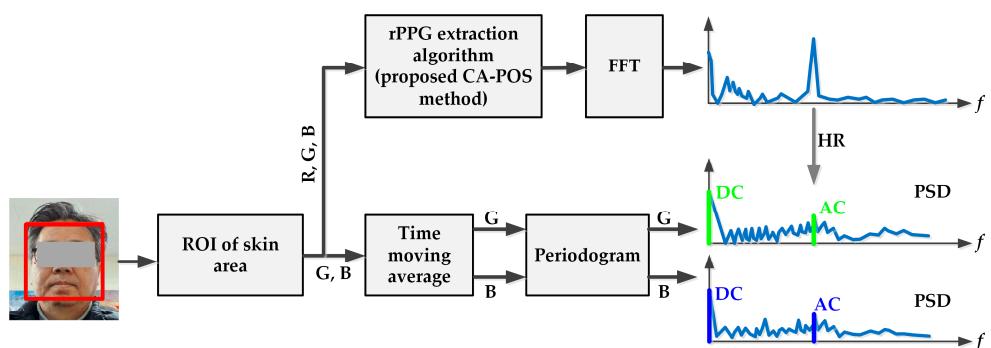


**Figure 3.** Filter sequence for HR extraction.

## 2.2. SpO<sub>2</sub> Detection

Blood contains hemoglobin, which is a compound of iron-containing heme and the protein globin. Oxidized hemoglobin reflects red light while significantly absorbing green light. For this reason, changes in the amount of oxygen in the blood can be modeled using the reflected intensity of the green channel. The PPG amplitude differs according to the wavelength (see Supplementary Figure S5). The noncontact oxygen saturation estimation method proposed in this paper uses the ratio of the green light absorption characteristics to the red light reflection characteristics of oxidized hemoglobin.

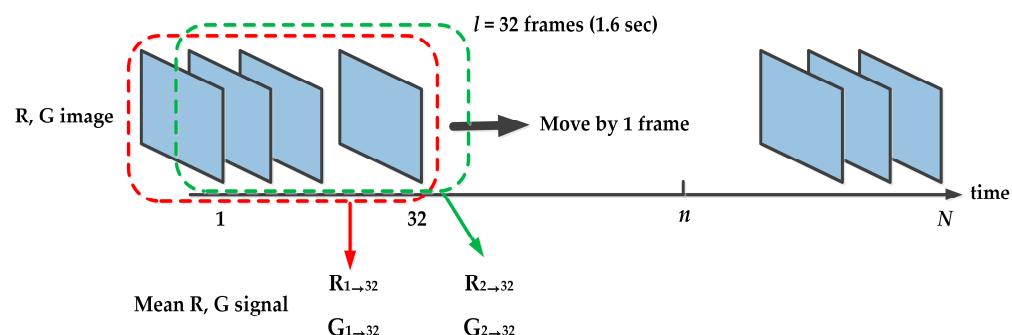
The proposed process for the noncontact measurement of the SpO<sub>2</sub> using a smartphone camera is shown in Figure 4. The SpO<sub>2</sub> measurement, using the estimated HR with high reliability, was measured simultaneously to the HR measurement. Only two RGB channels (R and G) were used for the SpO<sub>2</sub> measurement.



**Figure 4.** Proposed smartphone camera-based SpO<sub>2</sub> measurement process.

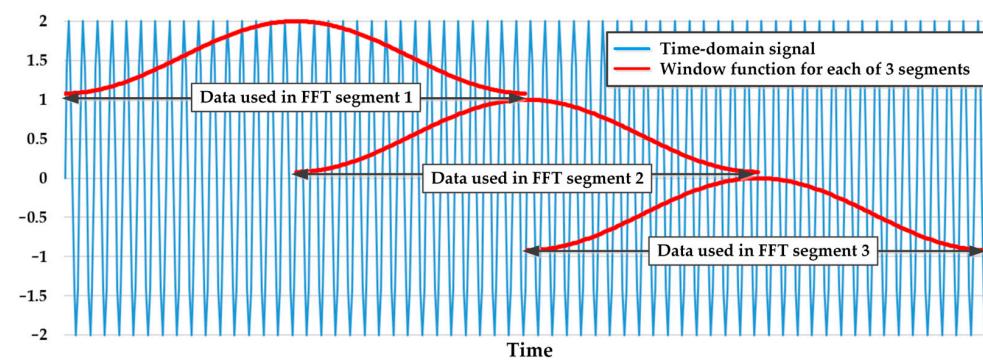
The process of extracting the R, G, and B time signals from the image obtained using a smartphone camera was the same as that for HR detection. Similarly, the ROI was detected using a face-tracking algorithm. The ROI only included the skin, including that of the face. Subsequently, the RGB time signals for the detected skin area were calculated by averaging the signal from each of the three RGB channels.

The R and G signals contained temporal noise such as lighting flicker; therefore, time moving average processing was performed to remove it, as shown in Figure 5, extracting the noise-containing R and G signals by moving one frame at a time for 32 frames (1.6 s). In the figure,  $R_{1 \rightarrow 32}$  indicates the result of averaging the R signal from the 1st frame to the 32nd frame.



**Figure 5.** Time moving average processing procedure for noise removal.

Despite implementing time moving average processing, the noise was not sufficiently removed from the R and G signals; therefore, a power spectral density (PSD) analysis was performed on these signals using a periodogram. A periodogram is primarily used for the frequency analysis of signals containing random noise. To obtain the periodogram, as shown in Figure 6, the time signal is overlapped and separated into N pieces using a window function, such as a Hamming window.



**Figure 6.** Time division to obtain periodogram.

A periodogram was obtained by averaging the PSDs of each temporally separated signal. Because the PSD can be obtained by using an FFT and squaring the time signal, the corresponding periodogram can be expressed as follows:

$$\frac{1}{N} \sum_{i=1}^N |FFT \text{ amplitue}(i)|^2 \quad (3)$$

The PSD of R and G obtained from the periodogram indicated a peak at the HR frequency. Because a single PSD includes random frequency components, determining a clear peak frequency is difficult. However, if N PSDs are averaged using a periodogram, then the noise frequency components cancel each other out, and the HR frequency component

can be maintained. In addition, because the HR frequency was extracted using a highly reliable HR detection algorithm such as the POS algorithm, the corresponding PSD value was used as the AC value, while the PSD value for a frequency of 0 was used as the DC value. Therefore, at the two wavelengths used to obtain the SpO<sub>2</sub>, the absorption factor K<sub>a</sub> could be estimated as follows:

$$K_a = \frac{\mu(K_R)}{\mu(K_G)} = \frac{AC_{\lambda_G}/DC_{\lambda_G}}{AC_{\lambda_G}/DC_{\lambda_G}} \quad (4)$$

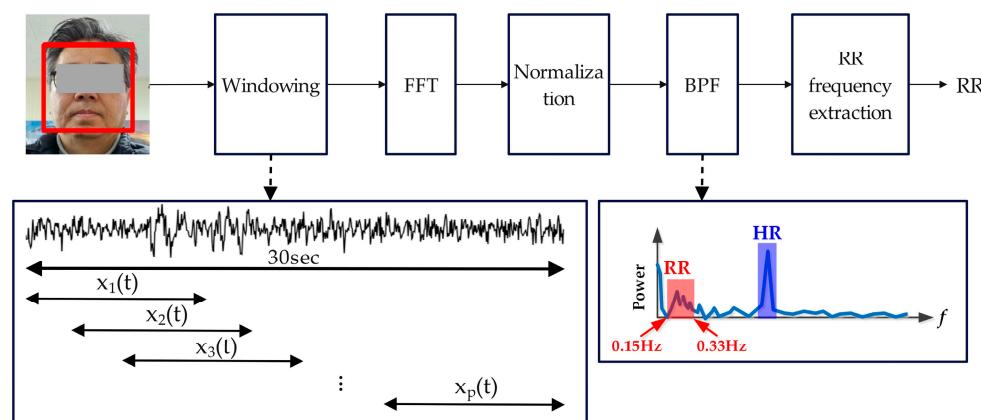
The expression for the calculation of the SpO<sub>2</sub> using the formula above was as follows:

$$\text{SpO}_2 = A - BK_a, \quad (5)$$

where A and B are constant values obtained empirically obtained through experiments.

### 2.3. RR Detection

Figure 7 shows the RR detection process using a smartphone camera. In the windowing process, an rPPG signal over 30 s was divided into multiple rPPGs of 7 s in length with a step size of 1 s. Each rPPG was subjected to FFT processing, and the FFT signals of the rPPGs were normalized. The normalized frequency spectrum was filtered by the respiratory frequency (0.15~0.33 Hz), which was determined using the maximum value in the normalized average frequency spectrum of an individual rPPG.



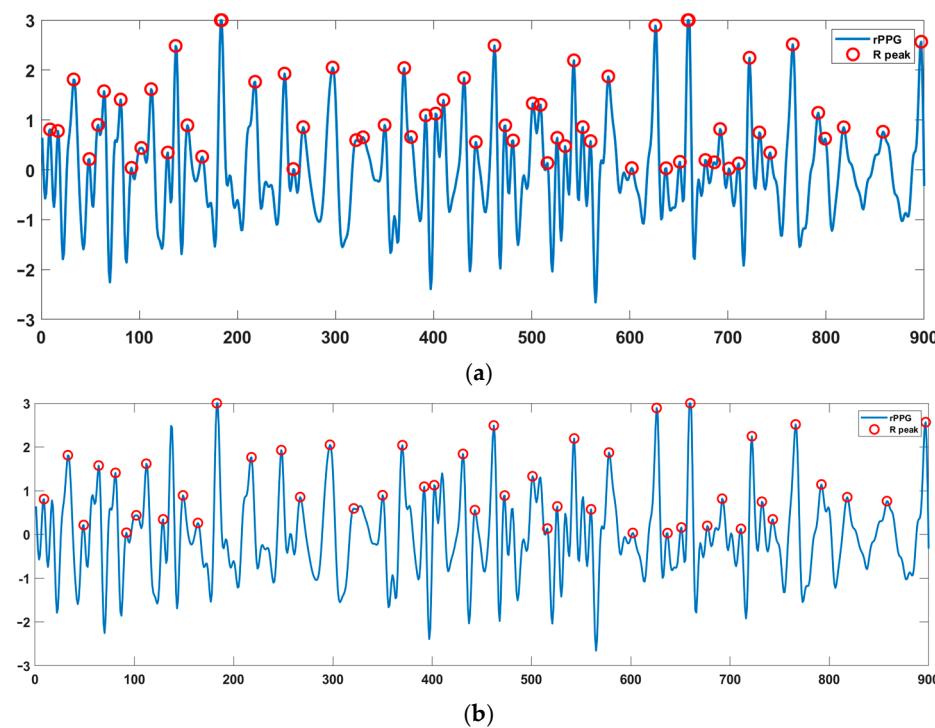
**Figure 7.** Proposed smartphone camera-based RR measurement process. The blue line represents the power spectrum of the rPPG signal.

We built a reference RR detection experimental environment to facilitate a comparison with the smartphone-based noncontact RR detection results (see Supplementary Figure S6). A NUL-236 recording sensor from NEULOG, Brackenheim, Germany, was used to measure the reference RR. While the test subject was breathing, a respiratory waveform was displayed based on the measurements of a pressure sensor inside the device, and the RR was calculated using the number of peak values in the respiratory waveform over a specific duration.

### 2.4. Heart Rate Variability and Stress Index

As shown in Figure 8, the rPPG was 30 s in length at 30 Hz (30 points per sec) and consisted of 900 points. By detecting the R peak in the rPPG, we were able to calculate the SDNN (standard deviation of NN) as a representative value of the HR variability. Additionally, the stress index was calculated using the HR, SDNN, and SpO<sub>2</sub> and ranged from 0 to 10. In the rPPG, the highest center point for a 1 × 5 filter size was identified as the

R peak. In this case, the center point also needed to be higher than the average of the rPPG signal. Additionally, two adjacent R peaks were considered one peak, and for this purpose, the second of these two peaks was excluded. The range of adjacent points was set to 10.



**Figure 8.** R peaks detected in rPPG signal. (a) 61 R peaks including adjacent R peaks. (b) 47 R peaks excluding adjacent R peaks.

Table 4 shows the stress scores (SSs) for the HR, SDNN, and SpO<sub>2</sub>. The higher the stress score of each bio-signal, the more severe the stress. Each stress score was proportional to the HR and inversely proportional to the SDNN and SpO<sub>2</sub>. By combining these, the final stress index (SI) was calculated as follows.

$$SI = \text{Round}((SS_1 + SS_2)/2 + SS_3) \quad (6)$$

**Table 4.** Stress scores based on (a) HR, (b) SDNN, and (c) SpO<sub>2</sub>.

HR Range	SS <sub>1</sub>	SDNN Range	SS <sub>2</sub>	SpO <sub>2</sub> Range	SS <sub>3</sub>
HR < 65	0	SDNN < 30	10	SpO <sub>2</sub> < 90	4
65 ≤ HR < 70	1	30 ≤ SDNN < 40	9	90 ≤ SpO <sub>2</sub> < 92	3
70 ≤ HR < 75	2	40 ≤ SDNN < 45	8	93 ≤ SpO <sub>2</sub> < 95	2
75 ≤ HR < 80	3	45 ≤ SDNN < 50	7	95 ≤ SpO <sub>2</sub> < 97	1
80 ≤ HR < 85	4	50 ≤ SDNN < 55	6	SpO <sub>2</sub> > 97	0
85 ≤ HR < 90	5	55 ≤ SDNN < 60	5		
90 ≤ HR < 95	6	60 ≤ SDNN < 65	4		
95 ≤ HR < 100	7	65 ≤ SDNN < 70	3		
100 ≤ HR < 105	8	70 ≤ SDNN < 80	2		
105 ≤ HR < 110	9	80 ≤ SDNN < 90	1		
110 ≤ HR	10	SDNN > 90	0		

## 2.5. Developed Software

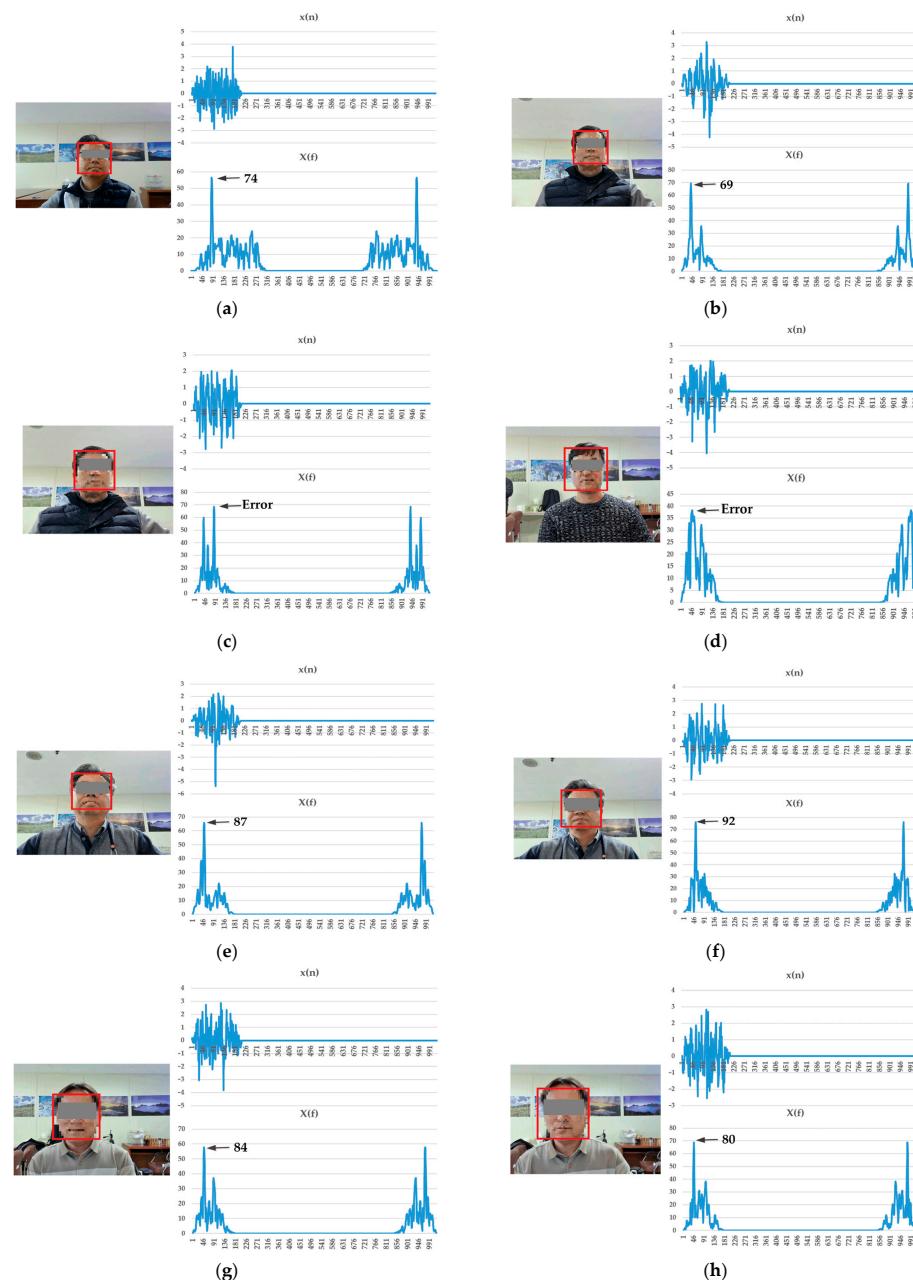
Software for noncontact bio-signal detection using a smartphone camera was developed in this study (see Supplementary Figure S7). It detects the HR and RR using images

taken with the camera and contains functions for selecting source images, setting the ROI, calculating the average values for ROIs, and detecting HRs and RRs.

### 3. Results

#### 3.1. HR Detection Results

We analyzed the HR detection results for the facial ROI in the image captured by the smartphone camera as shown in Figure 9. The upper signal,  $x(n)$ , in each subfigure displays the average value of all the pixels in the ROI on the time axis, and  $X(f)$  was obtained using zero padding and the FFT. As explained above, a BPF,  $H(f)$ , with a passband of 0.7 to 4.0 Hz and a Hamming window was used to detect the HR frequency band. In addition, the lower signal,  $Y(f)$ , in each subfigure is the spectrum resulting from BPF processing, where the HR frequency is the one with the peak value. Figure 9a–h contains the pulse frequency detected according to the facial angle.



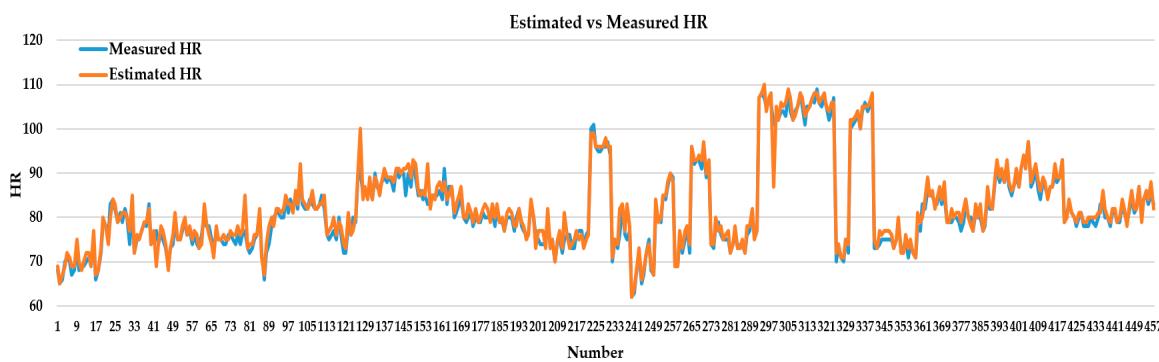
**Figure 9.** HR detection results for various face ROIs. The red boxes indicate the detected face areas.

Table 5 shows the noncontact HR detection results for 508 image sequences acquired using a smartphone camera under various conditions (face positions, illumination levels, distances from the camera, etc.). When the difference between the HR detected in a noncontact manner and the HR measured by the pulse oximeter was three or less, the HR was judged as normal; otherwise, it was judged as an error. In the experiments performed, the accuracy of the HRs obtained through noncontact detection and considered normal exceeded 88%. In addition, the remeasurement rate was approximately 10%. This implies that the noncontact HR detection method has a certain error rate and that the remeasurement algorithm proposed herein is extremely useful.

**Table 5.** Noncontact HR detection results.

	Normal	Remeasurement	Error	Sum
Detection result	447	51	10	508
Probability [%]	88.0	10.0	2.0	100

The correlations between the noncontact estimated HRs and the actual measured HRs for 457 video sequences, excluding HRs requiring remeasurement, are shown in Figures 10 and 11. In Figure 10, the orange line indicates the HR measured by the pulse oximeter, and the gray line indicates the HR estimated without contact using the smartphone camera. The actual measured HR and noncontact estimated HR were extremely similar. Figure 11a shows that the distribution of the actual measured HR vs. the noncontact estimated HR was linear. The error rate for significant differences between the noncontact estimated HRs was approximately 2%. Therefore, noncontact HR detection using a smartphone camera is suitable for use in real life. Figure 11b shows the Bland–Altman plots of the real HRs and those predicted using the proposed method. Bland–Altman plots are used to assess the level of agreement between two measurement methods. The difference between the two measurements is plotted against the mean of the two measurements, and a high density around the central ‘mean difference’ line within the ‘limits of agreement’ indicates strong agreement between the two schemes. The Bland–Altman plot revealed that the mean difference in the HR was  $-0.94$ , with a confidence interval of  $(-4.09, 2.56)$ . In the plot, the difference and mean results were extremely close together and appeared to overlap.



**Figure 10.** Noncontact HR detection results for image sequences excluding those requiring remeasurement (orange line: actual measured value; blue line: noncontact estimated value).

In order to quantitatively compare the performance of the proposed HR detection method and existing HR detection methods, the mean absolute error (MAE) and root mean

square error (RMSE) of the  $i$ -th actual HR,  $HR_i^{ref}$ , and the HR measured in N experiments,  $HR_i$ , were calculated as follows [20].

$$\text{MAE} = \frac{1}{N} \sum_i |HR_i - HR_i^{ref}| \quad (7)$$

$$\text{RMSE} = \frac{1}{N} \sqrt{\sum_i (HR_i - HR_i^{ref})^2} \quad (8)$$

The percentage of epochs (PEs) required for the HR estimation error to be less than 3.5 BPM was calculated as follows.

$$PE_{3.5} = \frac{1}{N} \left\{ i : |HR_i - HR_i^{ref}| < 3.5 \text{ BPM} \right\} \quad (9)$$

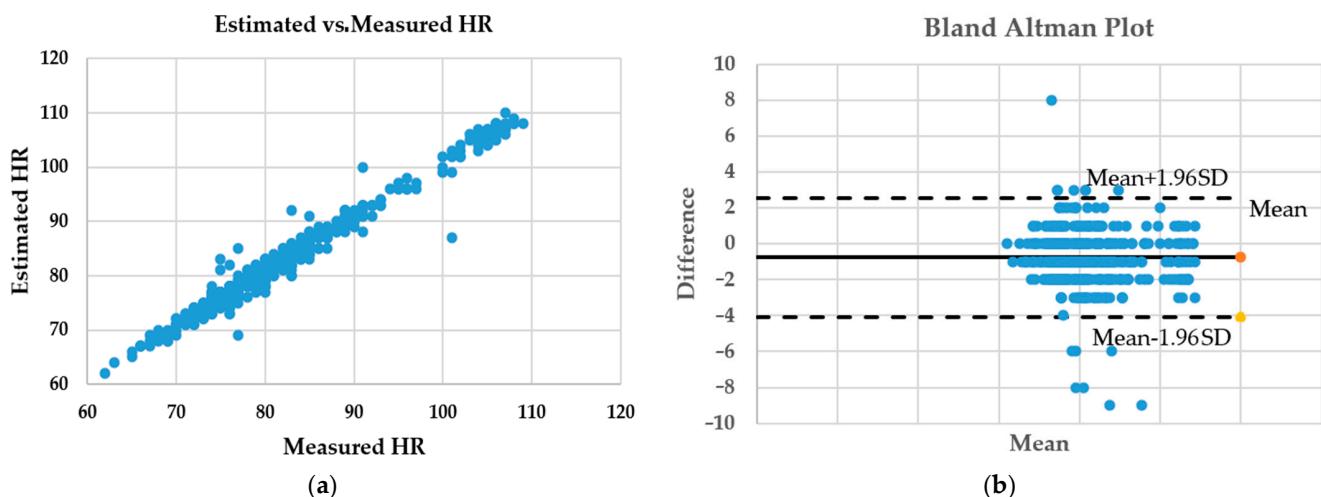
In addition, to evaluate the performance of the rPPG signal, the signal-to-noise ratio (SNR) was defined as follows.

$$\text{SNR} = \frac{1}{N} \sum_i 10 \log_{10} \frac{\sum_{f=40 \text{ BPM}}^{240 \text{ BPM}} (\hat{S}_i(f))^2 U_i(f)}{\sum_{f=40 \text{ BPM}}^{240 \text{ BPM}} (\hat{S}_i(f))^2 (1 - U_i(f))} \quad (10)$$

where  $\hat{S}_i(f)$  is the spectrum of the  $i$ th rPPG, and  $U_i(f)$  is defined according to whether the frequency component  $f$  is a signal ( $U_i(f) = 1$ ) or noise ( $U_i(f) = 0$ ).

$$U_i(f) = \begin{cases} 1 & \text{if } |f - HR_i^{ref}| \leq \Delta f \text{ or } |f - HR_i^{ref}| \leq 2\Delta f \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where  $\Delta f \approx 2.93 \text{ BPM}$ .



**Figure 11.** (a) Linear relationship between and (b) Bland–Altman plot of measured HRs and HRs estimated based on image sequences, excluding those requiring remeasurement.

Table 6 shows the MAE and other performance measures for the rPPGs estimated using the seven HR detection methods. For all the methods, RGB values were detected in facial ROIs using the AdaBoost method, and non-face regions were excluded from the calculations. Wavelet filtering was applied in all the methods except our proposed one. Regarding the four evaluation metrics for the extracted rPPG, the proposed CA-POS method and the 3D CNN method showed excellent results overall. G, GRD, AGRD, and ICA methods showed poor performance for the image sequences under various photography conditions used in this study. The 3D CNN method showed better performance in MAE

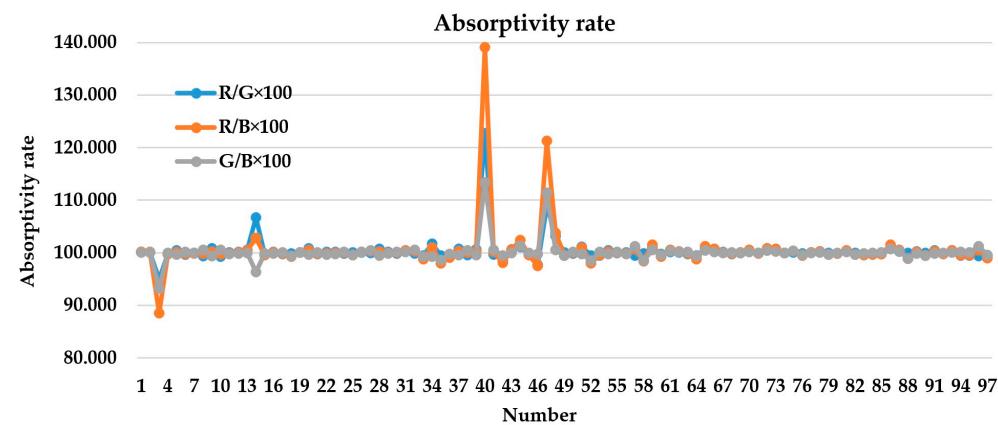
and RMSE BPM compared to CHROM, POS, LCTC, and the proposed method, but showed common performance in PE and SNR. The HR detection accuracy of the proposed method was over 88% and the mean difference in the HR was  $-0.94$ , with a confidence interval of  $(-4.09, 2.56)$  determined using a Bland–Altman plot. The mean errors for the SpO<sub>2</sub> and RR were 0.17 and 0.42, and their confidence intervals were  $(-2.96, 3.30)$  and  $(-2.92, 3.75)$ , respectively.

**Table 6.** Quantitative performance comparison of seven HR detection algorithms.

iPPG Extraction	MAE BPM	RMSE BPM	PE <sub>3.5</sub> , %	SNR, dB
G [5]	4.89	6.93	55	-2.89
GRD [5]	3.37	4.83	69	-0.73
AGRД [6]	3.89	5.16	63	-1.09
ICA [3]	3.12	4.67	67	-0.52
CHROM [9]	2.39	3.78	78	-0.16
POS [10]	2.28	3.51	84	0.28
LCTC [11]	2.23	3.42	80	0.25
3D CNN [12]	2.02	3.29	86	0.23
CA-POS (Proposed)	2.13	3.38	82	0.27

### 3.2. SpO<sub>2</sub> Detection Results

The absorptivity ( $K_a$ ) at the two wavelengths used to obtain the SpO<sub>2</sub> is shown in Figure 12. Most of the absorptance values were approximately 1, and the change in the absorptivity (green line) of G for B was the least significant. In the experiment, we assumed that the change in the absorption rate would be slight because it was being measured for normal people with insignificant changes in their SpO<sub>2</sub>. Therefore, to measure the SpO<sub>2</sub> more accurately, the error should be reduced by using the absorptivity of G with respect to B.

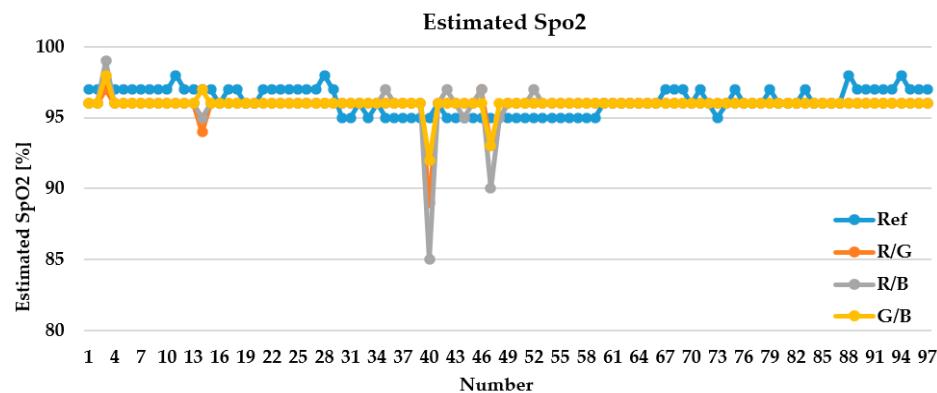


**Figure 12.** Absorptivity ( $K_a$ ) at two wavelengths used to obtain SpO<sub>2</sub>.

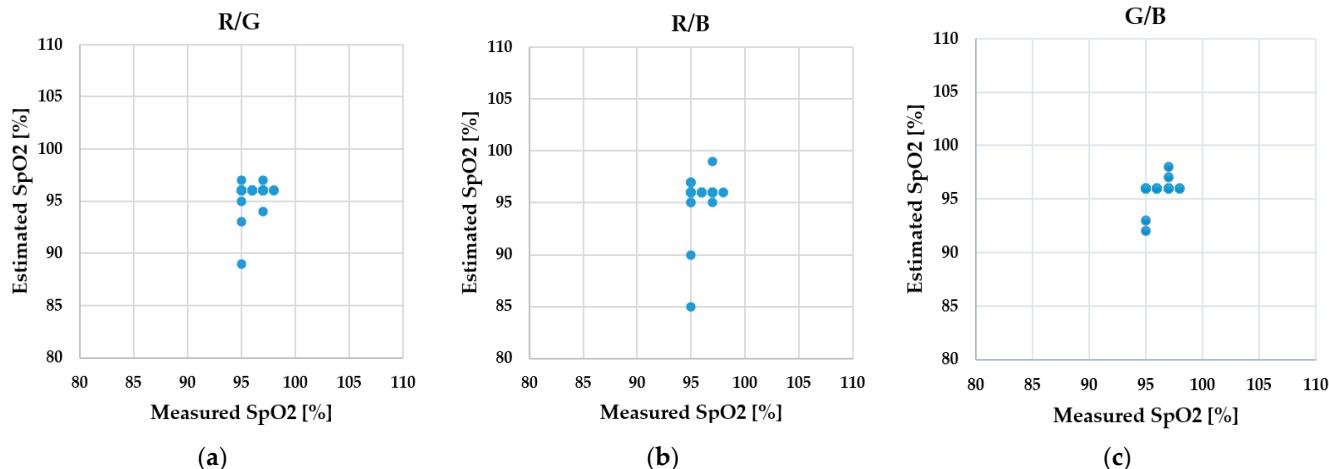
For the image sequence used in the experiment, Figure 13 shows the SpO<sub>2</sub> detection results based on the absorption rate for the two wavelengths, as well as the constants  $A = 125$  and  $B = 29$ . These constants are related to the absorption rate from the SpO<sub>2</sub> calculation expressed by Equation (5). The distribution of the SpO<sub>2</sub> detection results based on the absorption of each wavelength is shown in Figure 14, and the estimate using the absorption rate of G for B was similar to the actual measured value.

Figure 15 shows the MSE of the measured SpO<sub>2</sub> and the SpO<sub>2</sub> estimated by using the absorptivity at two wavelengths for 97 image sequences to obtain the absorptivity constants A and B. As shown in the figure, when  $A = 104$  and  $B = 8$ , the MSE for the absorption rate of G for B was the lowest. Because the noncontact detection value was similar to the actual

measurement value, we concluded that noncontact oxygen saturation detection technology using a smartphone camera is sufficient for utilization in noncontact bio-signal detection. However, since the experiment was only performed on healthy people with insignificant changes in their SpO<sub>2</sub>, additional research is required to determine the optimal absorption rate constants (A and B) for various healthy and patient groups.



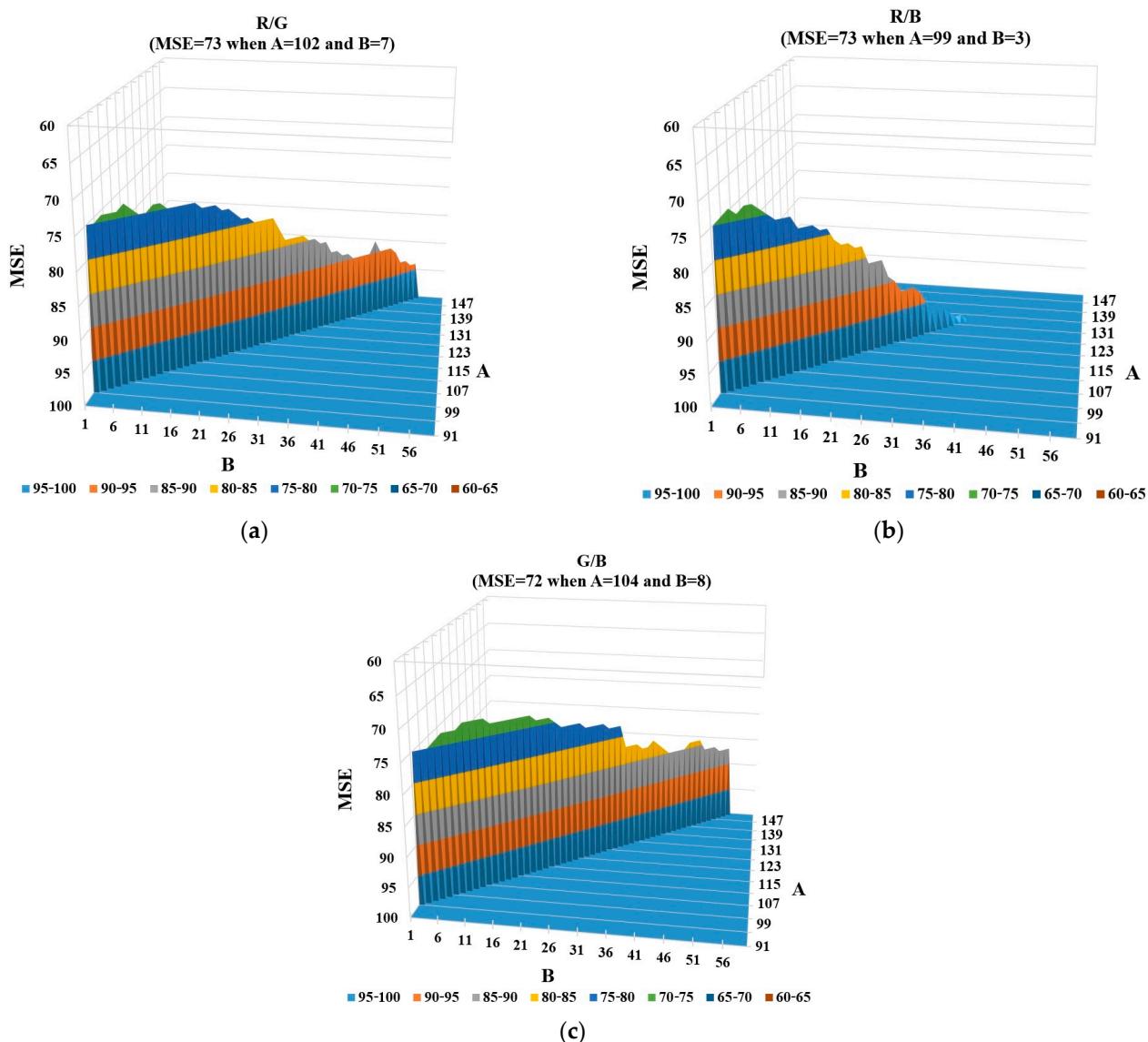
**Figure 13.** SpO<sub>2</sub> detection results for absorption at two wavelengths ( $A = 125$ ,  $B = 29$ ).



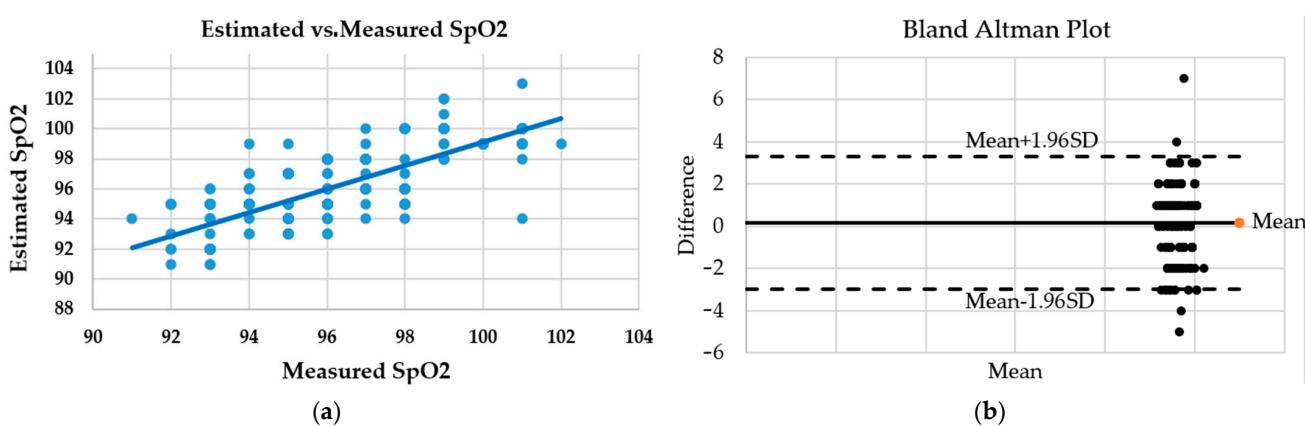
**Figure 14.** SpO<sub>2</sub> detection results estimated based on absorption of (a) R and G, (b) R for B, and (c) G for B.

Figures 16 and 17 show the SpO<sub>2</sub> and RR detection results obtained using the image sequence employed in HR detection. The Bland–Altman plot in Figure 16 shows that the mean error for the SpO<sub>2</sub> was 0.17 and the confidence interval was  $(-2.96, 3.30)$ . Additionally, the Bland–Altman plot in Figure 17 shows that the mean error for the RR was 0.42 and the confidence interval was  $(-2.92, 3.75)$ . It was confirmed that the noncontact estimated SpO<sub>2</sub> and RR had a linear relationship with the actual measured SpO<sub>2</sub> and RR.

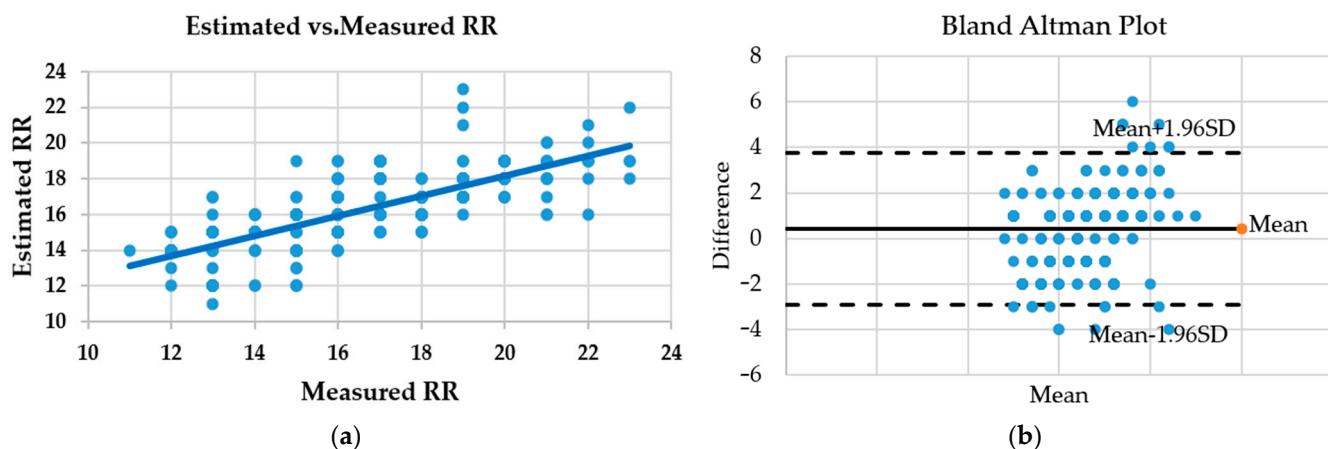
Table 7 presents a quantitative comparison of SpO<sub>2</sub> detection performance among existing methods and the proposed CA-POS method approach using MAE and RMSE as evaluation metrics. In terms of MAE, the proposed method achieved the lowest error of 0.86, outperforming the other methods. Regarding RMSE, the proposed method also demonstrated the lowest value of 1.23, indicating the most stable performance in terms of prediction variance. These results validate the effectiveness of the proposed method in improving SpO<sub>2</sub> estimation accuracy.



**Figure 15.** MSE of actual measured SpO<sub>2</sub> and SpO<sub>2</sub> obtained using estimated absorptance of (a) R and G, (b) R for B, and (c) G for B.



**Figure 16.** (a) Linear relationship between and (b) Bland–Altman plot of measured SpO<sub>2</sub> and SpO<sub>2</sub> estimated based on image sequences.



**Figure 17.** (a) Linear relationship between and (b) Bland–Altman plot of measured RR and RR estimated based on image sequences.

**Table 7.** Performance comparison of four SpO<sub>2</sub> detection algorithms.

Performance Criteria	Methods			
	Alharbi et al. [13]	Al-Naji et al. [14]	Akamatsu et al. [15]	CA-POS (Proposed)
MSE	1.18	1.25	0.98	0.86
RMAE	1.83	1.97	1.57	1.23

Table 8 provides a performance comparison of the proposed method and existing techniques for RR estimation, based on MAE and RMSE metrics. The proposed method achieved the lowest error values, with a MAE of 0.98 and an RMSE of 1.55. In contrast, the existing methods reported higher errors, respectively. These results confirm that the proposed method offers superior accuracy and stability in RR estimation compared to existing methods.

**Table 8.** Performance comparison of four RR detection algorithms.

Performance Criteria	Methods			
	Negishi et al. [16]	Zhao et al. [17]	Massaroni et al. [18]	CA-POS (Proposed)
MSE	1.12	1.34	1.21	0.98
RMAE	1.68	2.16	1.98	1.55

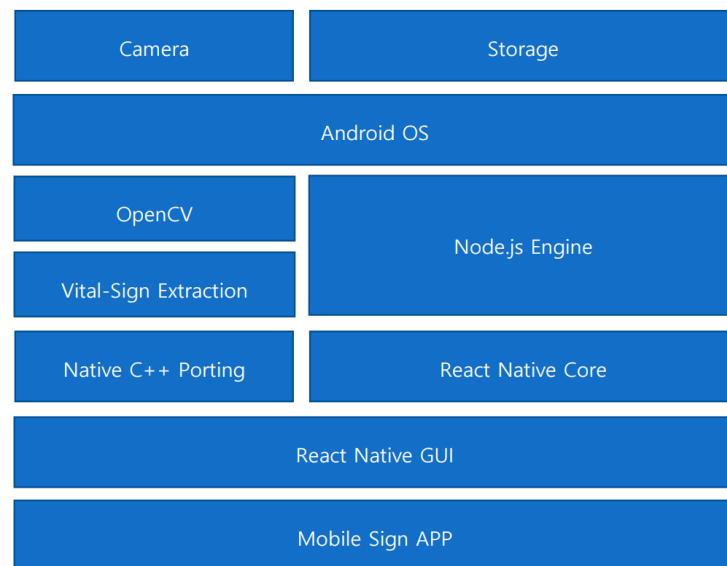
### 3.3. Mobile Application

The proposed noncontact bio-signal detection technique was implemented as a mobile app which measured, recorded, and analyzed the user's bio-signals using a smartphone camera and included a statistical data generation function. Considering the need for user convenience, interface development, and easy maintenance, React Native [21], a cross-platform development framework, was used for mobile application development rather than native platform development, as shown in Figure 18.

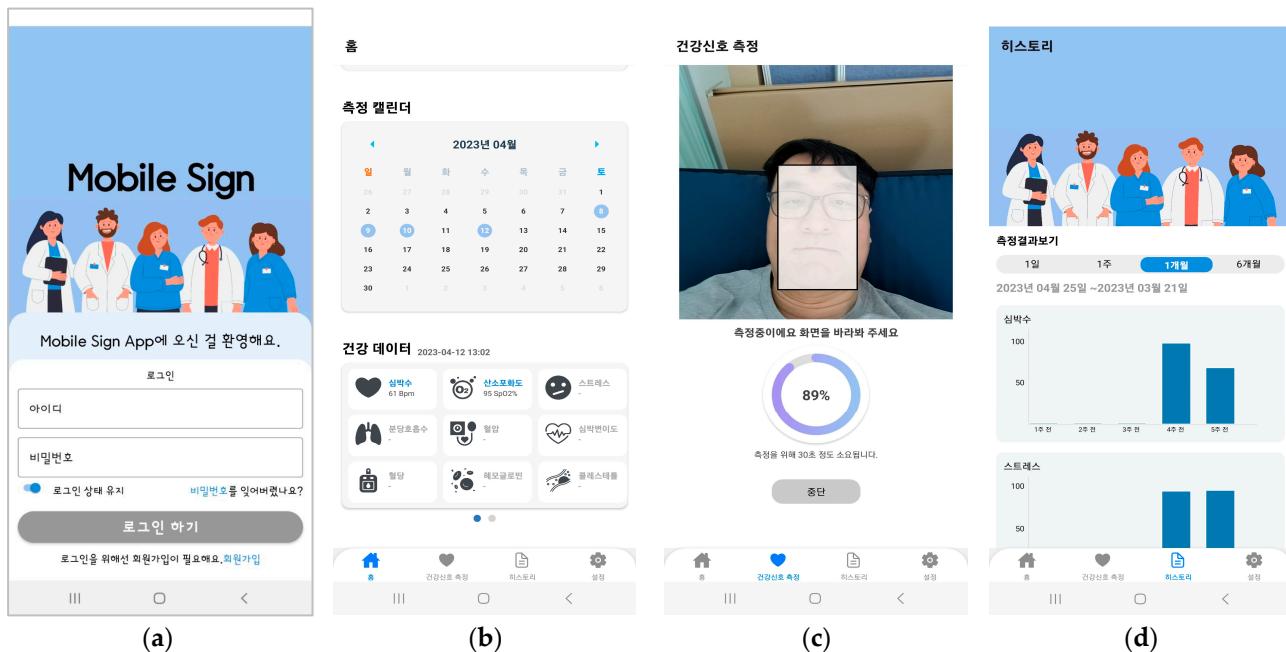
The user data management in the developed application was designed to be suitable for mobile storage-based data recording/management. The basic structure consists of a String-based data structure of Key: Value (Object) in JSON format, and will be designed in a form applicable to the Restful web service, non-normalized NoSQL databases, and SQL databases in the future.

The developed app is composed of login, home, measurement, and history screens, as shown in Figure 19. Currently, only the HR and SpO<sub>2</sub> functions are active. The home screen

consists of user information, a measurement calendar, and health data. When measured health signal data is added to the measurement calendar, a blue circle is displayed on the date. When a user selects the displayed date, the measured data appears in a health data tab. The health data can be scrolled through to the left and right, with up to five measurements per day, and five toggle buttons were created accordingly. The measurement screen's functions include (1) health data measurement, (2) result viewing, and (3) detailed result viewing.



**Figure 18.** System resource usage and configuration diagram of the developed app.



**Figure 19.** GUI of developed app: (a) login, (b) home, (c) measurement, and (d) history screens.

When viewing the history screen, the history over 1 day, 1 week, 1 month, and 6 months appears. The daily history view shows measured data as a blue circle on a calendar, just like the home screen calendar. When a date with data is selected, the user is taken to the measurement result screen. When the detailed measurement result view is selected, the user is taken to the detailed measurement result screen.

## 4. Discussion

### 4.1. Limitations of Proposed Method

The proposed method has several limitations, such as the lack of consideration for motion artifacts, the empirically fixed amplification scale for color difference signals, the limited accuracy of the SpO<sub>2</sub> estimation equation, the uncertainty regarding the appropriate frequency bands for bio-signal detection, and the insufficient size of the database. In this study, large subject movements in front of the camera were restricted, and the amplification scale for the color difference signals extracted from the detected facial region was determined empirically. In addition, the proposed SpO<sub>2</sub> estimation equation has not been thoroughly validated due to the limited database. Moreover, the frequency bands used for detecting HR, SpO<sub>2</sub>, and RR differ among researchers, and the optimal bands for various recording environments and video qualities have not yet been sufficiently investigated.

### 4.2. Challenges in Camera-Based Bio-Signal Detection

The first challenge is the quality of the image sequences captured by the camera. This issue is related to various camera-specific characteristics such as white point, brightness, gamma, focus, and the color correction matrix. To accurately detect facial color under different lighting conditions, these camera parameters must be properly considered. The second challenge concerns the recording environment and database limitations. Recently, non-contact bio-signal measurement methods using cameras have been actively researched, and several related mobile applications have been released [22]. However, due to issues such as illumination level, flicker, variations in skin tone across ethnicities, and user movement, the reliability of these methods remains relatively low.

### 4.3. Future Work

On the algorithmic side, we plan to address the automation of color difference signal amplification scaling and improve the accuracy of the SpO<sub>2</sub> estimation equation. We will also identify reliable frequency bands for HR and RR detection. To validate the general applicability of the proposed method, a comprehensive database will be constructed, encompassing various lighting conditions and diverse ethnic groups. In addition, to enhance the usability and accuracy of the currently developed mobile application, improvements will be made to both the app's user interface design and the measurement result analysis algorithms. Specifically, the convenience of reviewing bio-signal measurement history will be improved, and a food recommendation algorithm based on the user's current health status will be added. And, regarding the image sequence quality problem, we will focus on enhancing the removal of motion artifacts under diverse conditions by incorporating advanced deep learning architectures. In particular, the integration of deformable kernels, transformers, and diffusion-based models will be explored to improve robustness against complex motion and lighting variations [23]. Additionally, lightweight designs and training strategies across varied datasets will be considered to enable real-time performance and improved generalizability.

Future studies should focus on reducing the measurement bias in rPPG that arises from variations in skin tone across different ethnic groups. Darker skin tones, characterized by higher melanin concentrations, exhibit lower reflectance in the visible spectrum, which consequently degrades the reliability of physiological signal estimation. To overcome this limitation, future work will aim to incorporate large-scale training datasets that sufficiently represent a wide range of skin tones, employ color space transformation techniques for channel normalization, and implement skin-tone-aware deep learning compensation networks. In addition, multi-modal approaches using auxiliary channels such as IR, ther-

mal, or multispectral imaging will be investigated to structurally mitigate the optical bias introduced by skin pigmentation.

For the proposed method to be used in practical medical applications, it must demonstrate minimal differences compared to actual vital sign measurement devices. Additionally, to enhance user convenience, the method should be developed into applications compatible with smartphones, tablets, and small PCs. Furthermore, to accommodate elderly users, the applications will incorporate simplified interfaces and voice recognition technology.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/app15158509/s1>: Figure S1: Reflection model of rPPG signal; Figure S2: Mobile device camera-based HR measurement system; Figure S3: Hamming windows in (a) time and (b) frequency domains; Figure S4: HR detected by time interval; Figure S5: Relative PPG amplitude according to wavelength; Figure S6: Time moving average processing procedure for noise removal; Figure S7: Time divisions used to obtain periodogram.

**Author Contributions:** Conceptualization, T.W.B. and Y.C.K.; methodology, K.K.K. and T.W.B.; software, Y.C.K. and T.W.B.; validation, Y.C.K. and K.K.K.; formal analysis, Y.C.K. and T.W.B.; investigation, I.H.S.; resources, Y.C.K. and T.W.B.; data curation, I.H.S.; writing—original draft preparation, T.W.B.; writing—review and editing, T.W.B. and Y.C.K.; visualization, Y.C.K. and T.W.B.; supervision, K.K.K.; project administration, K.K.K.; funding acquisition, K.K.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by an Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government and the Gimhae Biomedical Industry promotion Agency. [25ZD1140, Regional Industry ICT Convergence Technology Advancement and Support Project in Daegu-Gyeongbuk (Medical); 25PD1210, Infrastructure Construction Project of Untact Medical Service Industry].

**Institutional Review Board Statement:** The present study's protocol was reviewed and approved by the Institutional Review Board of the Electronics and Telecommunications Research Institute (approval code N01-202304-01-006, 1 April 2023).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Acknowledgments:** We would like to thank Byeong-ik Kim (Electronic Advanced Development Team, AJIN Industrial Co., Ltd.) and Gwang-yong Kim (Daegu-Gyeongbuk Research Center, Electronics and Telecommunications Research Institute) for their assistance with data and funding acquisition.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Chen, X.; Cheng, J.; Song, R.; Liu, Y.; Ward, R.; Wang, Z.J. Video-based heart rate measurement: Recent advances and future prospects. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 3600–3615. [[CrossRef](#)]
- Poh, M.Z.; McDuff, D.J.; Picard, R.W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express* **2010**, *18*, 10762–10774. [[CrossRef](#)] [[PubMed](#)]
- Kwon, S.; Kim, H.; Park, K.S. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Diego, CA, USA, 28 August–1 September 2012; pp. 2174–2177.
- Chen, D.Y.; Wang, J.J.; Lin, K.Y.; Chang, H.H.; Wu, H.K.; Chen, Y.S.; Lee, S.Y. Image Sensor-based heart rate evaluation from face reflectance using Hilbert–Huang transform. *IEEE Sens. J.* **2015**, *15*, 618–627. [[CrossRef](#)]
- Lin, K.Y.; Chen, D.Y.; Tsai, W.J. Face-based heart rate signal decomposition and evaluation using multiple linear regression. *IEEE Sens. J.* **2016**, *16*, 1351–1360. [[CrossRef](#)]
- Feng, L.; Po, L.M.; Xu, X.; Li, Y.; Ma, R. Motion-resistant remote imaging photoplethysmography based on the optical properties of skin. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 879–891. [[CrossRef](#)]

7. Zhang, Y.; Dong, Z.; Zhang, K.; Shu, S.; Lu, F.; Chen, J. Illumination variation-resistant video-based heart rate monitoring using LAB color space. *Opt. Lasers Eng.* **2021**, *136*, 106328. [[CrossRef](#)]
8. Suh, K.H.; Lee, E.C. Contactless physiological signals extraction based on skin color magnification. *J. Electron. Imaging* **2017**, *26*, 063003. [[CrossRef](#)]
9. Haan, G.D.; Jeanne, V. Robust pulse rate from chrominance-based rPPG. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2878–2886. [[CrossRef](#)] [[PubMed](#)]
10. Wang, W.; Brinker, A.C.; Stuijk, S.; Haan, G. Algorithmic principles of remote-PPG. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 1479–1491. [[CrossRef](#)] [[PubMed](#)]
11. Chen, C.-C.; Lin, S.-X.; Jeong, H. Low-Complexity Timing Correction Methods for Heart Rate Estimation Using Remote Photoplethysmography. *Sensors* **2025**, *25*, 588. [[CrossRef](#)] [[PubMed](#)]
12. Sun, J.; Zhang, Z.; Liu, J.; Zhou, L.; Hu, S. Heart Rate Estimation Algorithm Integrating Long and Short-Term Temporal Features. *Mathematics* **2024**, *12*, 3444. [[CrossRef](#)]
13. Alharbi, S.; Hu, S.; Mulvaney, D.; Barrett, L.; Yan, L.; Blanos, P.; Elsahar, Y.; Adema, S. Oxygen saturation measurements from green and orange illuminations of multi-wavelength optoelectronic patch sensors. *Sensors* **2019**, *19*, 118. [[CrossRef](#)] [[PubMed](#)]
14. Al-Najai, A.; Khalid, G.A.; Mahdi, J.F.; Chahl, J. Non-contact SpO<sub>2</sub> prediction system based on a digital camera. *Appl. Sci.* **2021**, *11*, 4255. [[CrossRef](#)]
15. Akamatsu, Y.; Onishi, Y.; Imaoka, H. Blood oxygen saturation estimation from facial video via dc and ac components of spatio-temporal map. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
16. Negishi, T.; Abe, S.; Matsui, T.; Liu, H.; Kurosawa, M.; Kirimoto, T.; Sun, G. Contactless vital signs measurement system using RGB-thermal image sensors and its clinical screening test on patients with seasonal influenza. *Sensors* **2020**, *20*, 2171. [[CrossRef](#)] [[PubMed](#)]
17. Zhao, F.; Li, Y.M.; Qian, Y.; Tsien, J.Z. Remote measurements of heart and respiration rates for telemedicine. *PLoS ONE* **2013**, *8*, e71384. [[CrossRef](#)] [[PubMed](#)]
18. Massaroni, C.; Schena, E.; Silvestri, S.; Taffoni, F.; Merone, M. Comparison of two methods for estimating respiratory waveforms from videos without contact. In Proceedings of the IEEE International Symposium on Medical Measurement and Applications (MeMeA), Istanbul, Turkey, 26–28 June 2019.
19. Sun, G.; Suo, C.G.; Zhang, W.B. Face Gabor Feature Selection Based on AdaBoost. *Adv. Mater. Res.* **2013**, *694–697*, 1906–1909. [[CrossRef](#)]
20. Unakafav, A.M. Pulse rate estimation using imaging photoplethysmography: Generic framework and comparison of methods on a publicly available dataset. *Biomed. Phys. Eng. Express* **2018**, *4*, 045001. [[CrossRef](#)]
21. React Native. Available online: <https://reactnative.dev> (accessed on 15 April 2025).
22. New Technology Uses Cell Phone Camera. Cheeks to Take Vital Signs. Available online: <https://spectrumnews1.com/ca/lawest/health/2021/03/09/vital-signs-taken-using-camera-on-cell-phone-and-cheeks-only> (accessed on 25 April 2025).
23. Han, Z.; Li, L.; Jin, W.; Wang, X.; Jiao, G.; Liu, X.; Wang, H. Denoising and Motion Artifact Removal Using Deformable Kernel Prediction Neural Network for Color-Intensified CMOS. *Sensors* **2021**, *21*, 3891. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.