*Article*

# Integrating Remote Photoplethysmography and Machine Learning on Multimodal Dataset for Noninvasive Heart Rate Monitoring

Rinaldi Anwar Buyung [1], Alhadi Bustamam [1,2,*] and Muhammad Remzy Syah Ramazhan [1,2]

1   Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Indonesia,
    Depok 16424, Indonesia; rinaldi33@ui.ac.id (R.A.B.); muhammad.remzy@sci.ui.ac.id (M.R.S.R.)
2   Data Science Center (DSC), Faculty of Mathematics and Natural Science, Universitas Indonesia,
    Depok 16424, Indonesia
*   Correspondence: alhadi@sci.ui.ac.id

**Abstract:** Non-contact heart monitoring is crucial in advancing telemedicine, fitness tracking, and mass screening. Remote photoplethysmography (rPPG) is a non-contact technique to obtain information about heart pulse by analyzing the changes in the light intensity reflected or absorbed by the skin during the blood circulation cycle. However, this technique is sensitive to environmental lightning and different skin pigmentation, resulting in unreliable results. This research presents a multimodal approach to non-contact heart rate estimation by combining facial video and physical attributes, including age, gender, weight, height, and body mass index (BMI). For this purpose, we collected local datasets from 60 individuals containing a 1 min facial video and physical attributes such as age, gender, weight, and height, and we derived the BMI variable from the weight and height. We compare the performance of two machine learning models, support vector regression (SVR) and random forest regression on the multimodal dataset. The experimental results demonstrate that incorporating a multimodal approach enhances model performance, with the random forest model achieving superior results, yielding a mean absolute error (MAE) of 3.057 bpm, a root mean squared error (RMSE) of 10.532 bpm, and a mean absolute percentage error (MAPE) of 4.2% that outperforms the state-of-the-art rPPG methods. These findings highlight the potential for interpretable, non-contact, real-time heart rate measurement systems to contribute effectively to applications in telemedicine and mass screening.

**Keywords:** heart rate; machine learning; remote photoplethysmography; signal processing

## 1. Introduction

Heart rate is one of the most important vital signs observed to determine a person's health. This is because heart rate directly indicates cardiovascular diseases (CVDs) such as heart disease, arrhythmia, stroke, and high blood pressure. CVD is a noncommunicable condition that can lead to sudden death due to issues with the heart and blood vessels [1,2]. In addition, heart rate can also be an indicator of physical and emotional stress. The precious information provided by the heart rate makes it important to monitor. There are some tools for heart rate monitoring, such as stethoscopes, electrocardiograms, and pulse oximeters. However, those methods often involve uncomfortable physical contact with sensors, presenting challenges in continuous monitoring and real-time data collection [3–5].

The first non-contact heart rate monitoring techniques using optical sensors were invented by Hertzman in 1937, developing with the advancement of science and technology in medicine, known as photoplethysmography (PPG) [6]. PPG allows the measurement of the heart rate by placing a sensor directly on the skin, typically at areas like the fingertip or earlobe, to detect changes in the light absorbed or reflected by the skin due to blood volume pulsatile. With the development of light source technology and photo detectors in

the 1960s, more accurate and portable PPG devices were feasible. With further advances in electronics and signal processing in the 1970s, applications of PPG became more feasible in a clinical environment, as devices could now be smaller and more integrated.

In the early 2000s, researchers began exploring the potential of detecting PPG signals remotely using regular digital cameras which later provided the roots for remote photoplethysmography (rPPG). The main idea of rPPG is that skin color changes from the circulating blood flow were to be recorded and analyzed, mainly on facial regions. Initial attempts demonstrated that small changes in skin color due to blood flow could indeed be captured using webcams and camcorders, paving the way for remote, non-contact measurements. By using this technique, heart rate monitoring can be carried out remotely and faster. One of the rPPG technique implication is mass screening for life insurance underwriting. By using this approach, the assessment of prospective policyholders is much faster and cheaper [7–9].

Different algorithms have been developed to perform heart rate calculations using the rPPG technique. The first algorithm developed is called Green's algorithm, presented by Verkruysse et al. [10] in 2008. This algorithm highlights the green channel to find changes in blood volume within cutaneous blood vessels, which can later be used to monitor heart rate. The reason for choosing the green channel is related to the favorable absorption characteristics of hemoglobin. However, it is susceptible to individual differences in skin pigmentation, which might result in changing the obtained signal because of the various amounts of light absorbed and reflected.

In 2010, Poh et al. [11] developed a method for calculating heart rate named Independent Component Analysis (ICA). This method is a statistical method used to separate overlapping signals into independent components. The use of ICA in rPPG is very important because it helps improve the accuracy of noninvasive heart rate monitoring via video, making it very useful in more convenient and easy health monitoring. However, ICA has a major weakness—the assumption that the source signals that want to be separated are completely independent of each other. In practice, especially in the rPPG technique, the signal is related to the heartbeat, and the noise is not always completely independent.

Therefore, the next method is proposed, namely chrominance-based (CHROM) rPPG signal extraction proposed by De Haan and Jeanne [12] in 2013. CHROM does not completely rely on the assumption of independence. Instead, CHROM utilizes color relationships in the video to separate the heartbeat signal from noise so that it is more robust to situations where the signal is not completely independent. Although more stable than ICA, CHROM can still be affected by drastic changes in lighting because these changes affect the chrominance values used. Therefore, this method was refined by research conducted by W. Wang et al. [13] in 2017, who proposed the algorithm plane orthogonal to skin (POS). Due to the linear combination of color signals in the POS representation, it is invariant to the changes in the lighting conditions and, correspondingly, reduces the intensity variations within an improved signal stability.

Recently, with the emergence of deep learning, many studies on noninvasive heart rate measurement are shifted towards deep learning models such as convolutional neural network (CNN) [14–17]. However, CNNs often work as black boxes, making it difficult to understand how the inputs lead to the outputs. The lack of interpretability may raise concerns about accountability and reliability in many applications, especially in healthcare. On the other hand, rPPG has an advantage over CNN in terms of interpretability since the theoretical background of rPPG is rooted in optics and physiology, not to mention that the rPPG mechanisms are straightforward, making it easier for practitioners to understand how the measure correlates with physiological signal [18].

Although progress has been made in rPPG for contactless heart rate measurement, the current methods still struggle with noise and other external influences, including lighting conditions and the skin pigmentation of the individual. Such challenges hint that the current methods need a more wholesome approach by incorporating machine learning integration and multiple data types to enhance robustness and accuracy. In the event of these gaps, our

research proposes incorporating integration machine learning and a multimodal approach: 1 min facial video data along with physical characteristics, such as age, gender, height, weight, and body mass index (BMI), to enhance heart rate estimation. Two machine learning algorithms, namely support vector regression (SVR) and random forest, will be utilized to model the data input to predict the heart rate. Ablation experiments will be employed by removing or replacing each modality to investigate the impact on the model prediction. Our work will contribute to the integration of robust heart rate measurement for more personalized estimation, considering variability between individuals.

## 2. Methodology

### 2.1. Hypothesis

A study published by Quer et al. [19] utilized wearable fitness trackers to investigate age, body mass index, and sex-related differences in resting heart rate among 92,457 adults. It was concluded that the association of age and body mass index with RHR were statistically significant but small. Overall, the RHR tends to decrease with increasing age and was slightly higher with increased BMI. Similarly, two studies conducted by Ehrenwald et al. [20] and Gonzales et al. [21] studied the temporal changes in RHR. It has been observed that in general, the RHR decreases as age increases and women have a higher RHR compared to men. Furthermore, changes in BMI also affect the RHR; an increase in body mass results in an increased heart rate. Consequently, based on this study, we hypothesized that the inclusion of the rPPG technique and physical attributes, such as age, gender, weight, height, and BMI, may enhance the measurement system in its performance and reliability aspects.

### 2.2. Remote Photoplethysmography

Remote photoplethysmography (rPPG) is an innovative technique that allows measuring the heart rate noninvasively and without direct contact with the skin. The basic principle involves using light emitted onto the surface of the skin and detecting changes in the intensity of light reflected or absorbed by the skin during the blood circulation cycle [22]. The resulting signal, known as blood volume pulse, describes changes in blood vessel volume over time and allows the measurement of heart rate. The skin color measured by the camera is a combination of the intensity and spectrum of the light source, the intrinsic skin color, and the sensitivity of the camera's color channels. Temporal changes in skin color are proportional to the magnitude of lighting intensity that would be influenced by specular variations caused by movement and subtle color changes due to blood circulation. In rPPG, two components contribute to the observed color variations on the skin surface: specular and diffuse reflections.

Specular reflection refers to light reflection due to a smooth surface, such as the skin. This is the result of the direct reflection of light from the skin without any scattering. For many cases, it may be the most dominant component in observed color variations and overshadows other components. However, this does not contain information regarding heart rate, so it needs to be rejected or minimized in the rPPG technique. On the other hand, diffuse reflection may be said to deal with skin tissue absorption and light scattering. This is influenced by factors such as the hemoglobin and melanin content of the skin [23]. Diffuse reflection contains information related to changes in blood volume in the vessels around the skin. This component is of interest in rPPG because it can be used to extract signals related to changes in blood volume [24]. Figure 1 shows an illustration of the use of light emitted into the skin in the rPPG technique.
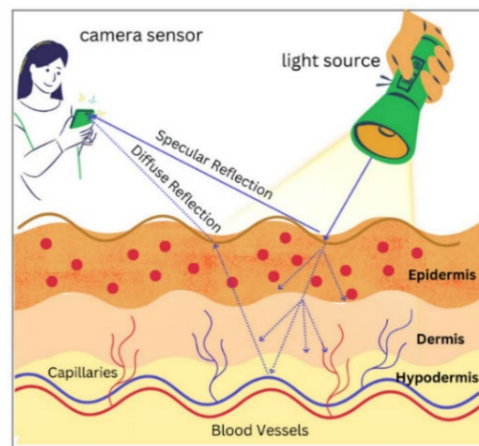
**Figure 1.** Illustration of remote photoplethysmography.

To extract heartbeat signals from observed color variations, the rPPG method focuses on isolating and analyzing diffuse reflection components while minimizing the impact of specular reflections. For this purpose, the Plane-Orthogonal-to-Skin (POS) approach put out by [13] is the methodology employed in this study. The main idea of the POS is that the video sequence captures a visible skin region, where color variations correspond to heartbeat-related blood flow changes. The POS algorithm aims to account for variations in skin tone by projecting the RGB signal onto a plane orthogonal to the temporally normalized skin tone direction when given a 15 s facial video where each frame's RGB values are averaged to obtain a single-color intensity value per frame, representing the overall color change over time. This results in a temporal array of $R(n)$, $G(n)$, and $B(n)$ where $n$ represents frame index. The array is then projected using a matrix $p_x = \begin{pmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{pmatrix}$, designed to minimize the influence of skin tone on pulse extraction. This guarantees that there are as many beat components as possible in the final rPPG signal. After that, the signal's component will be separated to calculate heart rate representation. The algorithm of POS is formulated in Algorithm 1 as follows:

---

**Algorithm 1:** Plane orthogonal to skin

---

Assumption: Video duration 15 s; Frame rate 30 fps; total frame 450 frames
Input: Video contains $N$ frame
Output: rPPG signal

1:      $H$ = zeros $(l, N)$, $l = 48$ (30 fps camera)
2:      for $n = 1, 2, \ldots, N$ do:
3:          $C(n) = [R(n), G(n), B(n)]^T$
4:          if $m = n - l + 1 > 0$ then
5:              $C_n^i = \frac{C_{m\to n}^i}{\mu \, C_{m\to n}^i}$
6:              $S = \begin{pmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{pmatrix} \cdot C_n$
7:              $S_1 = S[1]$ # The first row of $S$
8:              $S_2 = S[2]$ # The second row of $S$
9:              $h = S_1 + \frac{\sigma(S_1)}{\sigma(S_2)} \cdot S_2$
10:            $H_{m\to n} = H_{m\to n} + (h - \mu(h))$
11:         end if
12:      end for
13:      return $H$

---

This algorithm has the benefit of being less constrictive regarding the degree of distortion shown in the video and more resilient against noise. It is appropriate for a range of applications since it can function well in both fixed and moving scenarios.

### 2.3. Signal Filtering

In the actual world, noise frequently taints signals. There are many different types of noise, such as environmental interference, jitter in data acquisition, and electromagnetic interference. To make the signal being measured or analyzed cleaner and more precise, noise can be reduced or eliminated using the filter process. Butterworth bandpass filter, which is thought to enhance the quality of rPPG signals, is one of the most popular filters [25,26]. According to Steiglitz [27], Butterworth filters are made to have a flat frequency response within the passband frequency range, which permits all frequencies to go through without distortion. The transfer function, a polynomial ratio of output to input, is a way the filter process response to input signals may be described. The mathematical expression for the transfer function is as follows:

$$H(s) = \frac{1}{s^2 + (2\sqrt{n} \cdot \pi \cdot f_c)s + N\pi^2 f_c^2} \tag{1}$$

where the transfer function $H(s)$ is the transfer characteristics, and $s$ is a complex variable function. The parameter $N$ indicates the filter order, whereas $f_c$ signifies the cutoff frequency. In this study, a Butterworth filter will use the cutoff parameters in the range 0.8–2.0 Hz and the order used is 4. The reason for choosing this interval is because the average resting human heart rate is in the interval 48–120 bpm [28]. A fourth-order Butterworth filter is chosen for several reasons. First, higher-order filters provide steeper roll-offs, which enhance the filter's ability to sharply distinguish between the passband and the stopband.

### 2.4. Fast Fourier Transform

After being filtered, the rPPG signals may now be used to obtain crucial information for heart rate computations. The Fourier transform may be used to analyze frequency by applying it to the rPPG signal. Signals in the time domain will be converted into the frequency domain using this technique. Consider a discrete time signal $x[t]$ of length $N$ ($t = 0, 1, 2, ..., N-1$). Discrete Fourier transform (DFT) from $x[t]$ can be defined as follows:

$$X[f] = \sum_{t=0}^{N-1} x[t]e^{-j\frac{2\pi kt}{N}} \tag{2}$$

where $X[f]$ is the $f_{th}$ frequency spectrum, $x[t]$ is time signal of $t_{th}$ sample, and $N$ is the signal length. However, $O(n^2)$ operations are needed for classical DFT, which can be quite slow for long signal lengths. This issue is resolved by the fast Fourier transform (FFT), which divides the DFT into smaller DFTs that can be computed more quickly. The Cooley–Tukey algorithm is the most well-known FFT algorithm. The FFT is split into two smaller pieces by this algorithm, which then processes each piece recursively [29].

### 2.5. Support Vector Machine

The support vector machine is one of the most powerful models of machine learning widely used for both classification and regression. SVMs which are used for regression tasks are called support vector regression (SVR). It has based its idea on finding the optimal decision boundaries or a hyperplane with maximum margin [30]. The main optimization problem of the hyperplane for regression involves the minimization of the sum of slack variables, both positive and negative, allowing deviations within a margin $\epsilon$ from the true

target value. Let $N$ be the total number of data points in the training set; then, the formula is as follows:

$$
\min_{w,b} \tfrac{1}{2}\|w\|^2 + C \sum_{n=1}^{N} (\xi_n + \hat{\xi}_n)
$$
$$
s.t\ t_n \leq y(x_n) + \epsilon + \xi_n
$$
$$
t_n \geq y(x_n) - \epsilon - \hat{\xi}_n \tag{3}
$$
$$
\xi_n,\ \hat{\xi}_n \geq 0
$$

The $w$ defines the weight vector that sets the hyperplane separating the classes in a higher-dimensional space. Therefore, $\|w\|^2$ is the squared norm of the weight vector, which represents the margin of separation between classes. Variable $C$ is a regularization parameter that controls the trade-off between maximizing margin and minimizing errors. Two slack variables $\xi_n$ and $\hat{\xi}_n$ are introduced to handle deviations above and below the margin. The Figure 2 below illustrates the hyperplane problem of the SVR and the role of each variable.
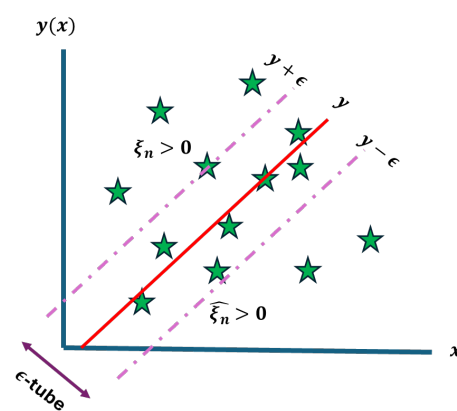


**Figure 2.** Hyperplane problem of the SVR.

However, Equation (3) is complicated to solve. Therefore, the dual form will be formulated using the Lagrange function and Karush–Kuhn–Tucker (KKT) condition. For regression problems, the dual of the hyperplane optimization method can be obtained using similar techniques. Therefore, the dual of regression hyperplane optimization is given as follows:

$$
\underset{a,\ \bar{a}}{\mathrm{argmax}} -\tfrac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} (a_n - \bar{a}_n)(a_m - \bar{a}_m) k(x_n, x_m) - \epsilon \sum_{n=1}^{N} a_n + \bar{a}_n
$$
$$
+ \sum_{n=1}^{N} (a_n + \bar{a}_n) t_n
$$
$$
s.t\ 0 \leq a_n \leq C,\ 0 \leq \bar{a}_n \leq C \tag{4}
$$
$$
\sum_{n=1}^{N} (a_n - \bar{a}_n) = 0
$$

where $a_n$ is associated with the upper constraint $t_n \leq y(x_n) + \epsilon + \xi_n$ and $\bar{a}_n$ is associated with the lower constraint $t_n \geq y(x_n) - \epsilon - \hat{\xi}_n$. Variable $k(x_n, x_m)$ is a kernel function. There are several kernel functions such as linear, polynomial, and RBF (Radial Basis Function) kernel [31].

The SVR model can be implemented for multimodal heart rate prediction by incorporating signal features derived from photoplethysmography alongside the physical information, such as age and body mass index, to improve accuracy. The SVR will build the best model based on the optimization of the hyperplane problem during the training process. The best model represented in mathematical function will be used to predict the heart rate based on the input. To enhance machine learning performance, the hyperparameters of the SVR need to be tuned. One of the most important hyperparameters to be

tuned is the kernel function. The kernel function converts data into the right format, hence allowing SVRs to identify the best boundaries. Another hyperparameter to be tuned is $C$ which is the balance between attempting to fit the training data with a minimal error and maximizing the margin of the decision boundary; a higher value of $C$ will make the considerations of correct classification more important. $\gamma$ is a factor associated with the influence of training data on the decision boundary, whereas $\epsilon$ is a margin within which no penalty is incurred in the training loss function [32].

*2.6. Random Forest Regression*

The random forest regression is an ensemble machine learning model that relies on the bagging approach: building multiple base models in parallel, and the final prediction is performed with an average vote, ensuring variance reduction by building models using several bootstrap datasets [33]. It consists of several fixed-size base models, which are usually classification and regression trees (CARTs). It builds binary trees, where each internal node splits the data into two child nodes based on a split criterion. The goal is to create a tree structure that maximizes the accuracy of predictions for new data points.

Given the training data $\{x_n, t_n\}$, $n = 1, \ldots, n$, where $t_n$ is a label. Let $Q$ be the set of data at a node, $\theta = (j, a)$ be the set of candidate feature $j$ and threshold value $a$, and $Q_{left}(\theta)$ and $Q_{right}(\theta)$ be the data in the left and right partitions of the threshold, defined as the following equation:

$$
\begin{aligned}
Q_{left}(\theta) &= \left\{ (x,t) \middle| x^{(j)} \le a \right\} \\
Q_{right}(\theta) &= Q - Q_{left}(\theta) \\
G(Q, \theta) &= \frac{N_{left}}{N} H\left(Q_{left}(\theta)\right) + \frac{N_{right}}{N} H\left(Q_{right}(\theta)\right)
\end{aligned}
\tag{5}
$$

where $N$ is the number of data points at the node, and $N_{left}$, $N_{right}$ are the numbers of data points in the left and right partitions, respectively. Therefore, selecting the feature and threshold that produces the purest partition at a node can be expressed as the following optimization problem:

$$
\theta^* = \min_\theta G(Q, \theta)
\tag{6}
$$

For regression tasks, the input is divided into several partitions called $R_m$. Each region contains a subset of the data points (observations) that fall within it. The goal is to find optimal partitions that minimize the error in each region. The node m represents a region $R_m$ with $N_m$ observations, and the mean of the target values within each region is $c_m$. The common criterion to minimize is the mean squared error $H(x_M)$ as shown in the following equations.

$$
\begin{aligned}
c_m &= \frac{1}{N_m} \sum_{i \in N_m} t_i \\
H(X_m) &= \frac{1}{N_m} \sum_{i \in N_m} (t_i - c_m)^2
\end{aligned}
\tag{7}
$$

Next, using the steps above, the random forest regression algorithm builds multiple decision trees by bootstrapping and aggregating the results. The algorithm to perform random forest regression is given in Algorithm 2 as follows:

---

**Algorithm 2:** Random forest regression

---

Input: training dataset, *n*: number of trees; *d*: tree depth; *m*: number of features at each node
Output: prediction of Random Forest model
1:    for *i* = 1, 2, . . ., *n* do:
2:        Bootstrap sampling of the dataset $D_i$
3:        Build decision tree:
4:        At each node, randomly sample m features
5:        Find the best split using the *m* subset
6:        Split the node and repeat until stopping criteria (max depth or minimum samples)
are met.
7:        Save the decision tree $T_i$
8:    end for
9:    For each tree $T_i$, get the predicted value from all trees
10:    return Average of the predictions from each tree

---

The random forest model can be implemented for multimodal heart rate prediction by incorporating signal features derived from photoplethysmography alongside the physical information, such as age and body mass index, to improve accuracy. Multiple decision trees will be built to model the data, and the final prediction will be calculated by averaging the result of each tree.

For random forests, there are three main hyperparameters that will be tuned to improve the model's performance. Firstly, the number of trees which controls how many trees are involved in the ensemble. Secondly, max depth is the setting that controls the depth of each tree, and this must be performed to avoid overfitting. The final hyperparameter, minimum samples per split denotes the minimal number of samples that are necessary to split an internal node, which prevents the trees from growing too complex [34]. The illustration of the random forest regression can be seen in Figure 3.
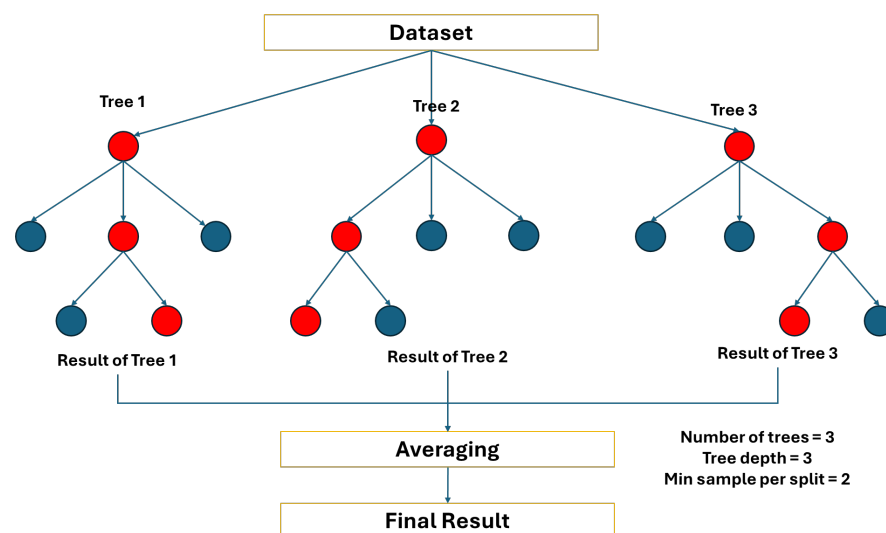


**Figure 3.** Random Forest Regression model.

## 2.7. Data Collection

In this experiment, a dataset with the facial video recordings of 60 individuals was utilized. Each video's length was standardized to 1 min, where the resolution and frame rate were set to be uniform at 480 × 640, and 30 fps, respectively. This resolution and frame rate are used to balance the performance and computational efficiency, leading to an accurate prediction but still feasible for real-time implementation. The data collection was conducted at the same place and with fixed lighting and subjects had to take off any facial accessories, such as eyeglasses. Together with video recording, there was manual

measurement using an oximeter to gather the actual value as the ground truth. The height in cm, weight in kg, age in years, and gender are also recorded in the dataset. To ensure privacy, we anonymized the subjects by masking facial features such as eyes, nose, and mouth to prevent unauthorized recognition. We also introduced variable body mass index (BMI) into the dataset. Figure 4 depict the characteristics of the dataset used in this study.
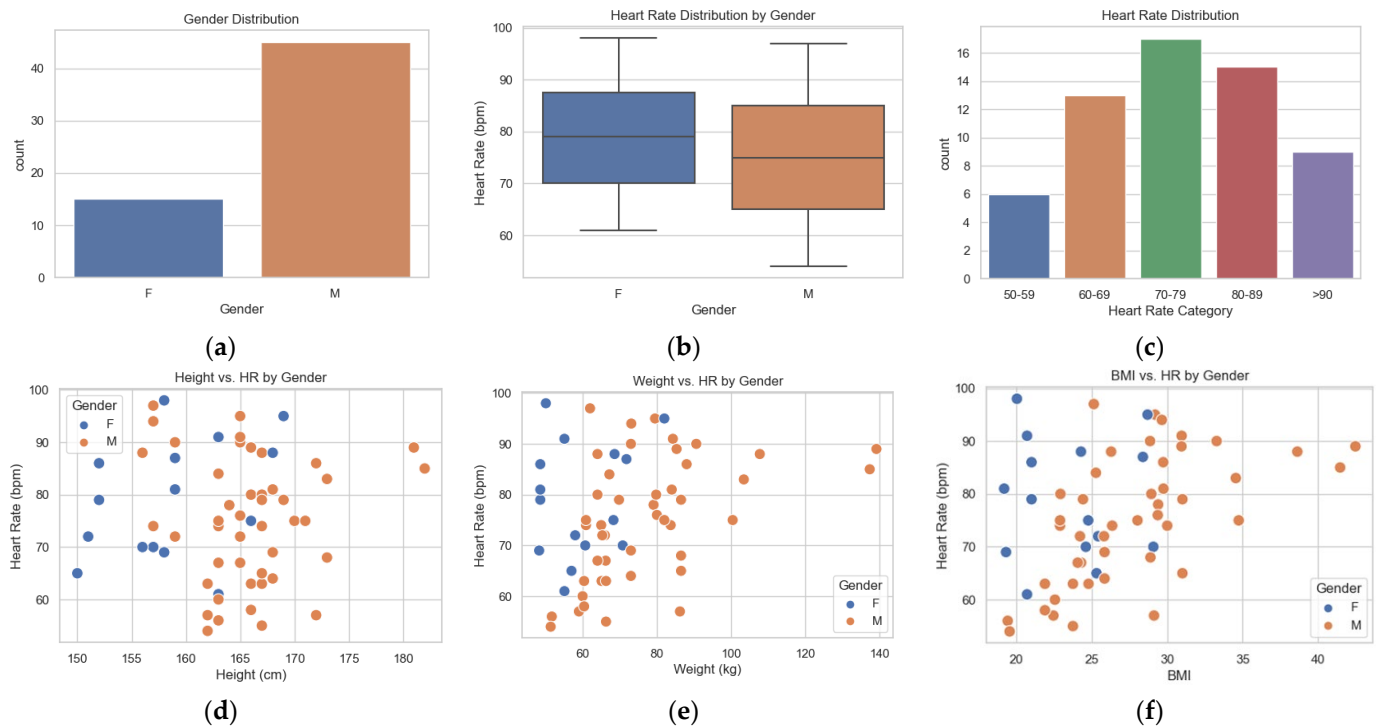


**Figure 4.** Dataset used in this study. (**a**) Gender distribution; (**b**) Heart rate distribution by gender; (**c**) Heart rate distribution by category; (**d**) Height vs. Heart Rate by Gender; (**e**) Weight vs. heart rate by gender; (**f**) BMI vs. heart rate by gender.

From Figure 4a, it can be concluded that from the 60 individuals, there are 45 males and 15 females in the dataset. Figure 4b shows the variations in heart rate between the males and females. It also shows that females generally present a higher heart rate compared to males, since their median rate is just above 75 beats per minute, while the median for males is just about 70 beats per minute. The interquartile range (IQR), which represents the middle 50% of the dataset, is wider for females, ranging approximately between 70 and 85 bpm. This indicates that there is a greater spread or deviation in the pulse rates of females. In contrast, the IQR for males varies between 65 and 80 beats per minute; this means that male heart rates are more huddled around the lower end of the spectrum. Figure 4c–f indicate that our dataset has a good diversity of heart rate, height, weight, and BMI, making this ideal for training machine learning.

## 2.8. Region of Interest Selection

Region of interest (RoI) selection in the rPPG technique involves selecting a specific area on a person's skin or face from which the rPPG signal will be extracted. The face was chosen as the RoI in the rPPG algorithm because it has several advantages over other body parts. First, the face has a high concentration of capillaries, making it easier to detect changes in blood flow. Second, faces have a low level of reflection, making it easier to separate PPG signals from environmental noise. Third, faces are easy to access and reach, making them suitable for real-time video capture applications [26,35]. For this purpose, we utilized Dlib facial landmarks that combine Histogram of Oriented Gradients (HOG) as the feature extraction and SVM as the key points classifier [36]. The RoI used in this research

is the facial skin without eyes, nose, and lips. For this purpose, the landmarks (36, 39, 40, and 41) for the left eye, (42, 45, 46, and 47) for the right eye, (32, 33, 34, 35, and 36) for the nose, and (49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, and 60) for the lips will be filled as black as shown in Figure 5.



**Figure 5.** Region of interest used in this study.

*2.9. Feature Selection*

In machine learning, signal feature selection is the process of identifying and selecting the most relevant features included in the signal data that contribute to an improvement in the performance of machine learning models. After several surveys in the literature [27,37–39], we decided to utilize signal features as follows:

- Maximum and minimum values of the signal: They are the basic features of a signal representing the maximum and minimum value of signal amplitude, respectively. From those values, a new feature called peak to peak, which is the difference between the maximum and minimum values of the signal, can be generated. Given that $x(t)$ represents the signal amplitude at time frame $t$, the maximum, minimum, and peak-to-peak values of the signal are represented as follows:

$$max_{value} = \max_t x(t)$$
$$min_{value} = \min_t x(t) \tag{8}$$
$$peak\ to\ peak = max_{value} - min_{value}$$

- Skewness and Kurtosis: Skewness is a measure of the asymmetry of the signal's amplitude distribution around the mean, while kurtosis is the existence of outliers in the signal's distribution. Given that $x(t)$ represents the signal amplitude at time frame $t$, and $\mu, \sigma$ are the mean and standard deviation of the signal, respectively, skewness and kurtosis of the signal are represented as follows:

$$x_{ske} = \frac{E\left((x(t)-\mu)^3\right)}{\sigma^3}$$
$$x_{kur} = \frac{E\left((x(t)-\mu)^4\right)}{\sigma^4} - 3 \tag{9}$$

- Entropy: It measures the complexity or randomness in the signal based on the probability distribution of its values. It is often used to assess the predictability of a signal. Given that $x(t)$ represents the signal amplitude at time frame $t$, to calculate entropy, the signal $x(t)$ needs to be discretized over a period of 30, matching the fps assumption used in this study. This process results in $x_i$ where $i = 1, 2, \ldots, 30$. Each $x_i$ covers a specific range that $x(t)$ can take. The entropy can be calculated as follows:

$$H(s) = -\sum_{i=1}^{30} p(x_i)\log_2 p(x_i) \tag{10}$$

where $p(x_i)$ represents the probability that $x(t)$ falls within the range specified by $x_i$. Higher entropy indicates a more complex or less predictable signal, while lower entropy suggests a more regular, predictable pattern.

- Zero crossings: Zero crossings are a count of how many times the signal crosses the horizontal axis—that is, changes from positive to negative or vice versa. This feature also gives some insight into the frequency content of the signal. Given that $x(t)$ represents the signal amplitude at time frame $t$, zero crossings can be formulated as follows:

$$Z = \sum_{i=1}^{n-1} \mathbb{I}(t,\ t+1)$$
$$\mathbb{I}(t,\ t+1) = \begin{cases} 1, & if\ x(t) \cdot x(t+1) < 0 \\ 0, & if\ x(t) \cdot x(t+1) > 0 \end{cases} \tag{11}$$

- Spectral centroid and spectral bandwidth: The spectral centroid is that point, considering the signal in the frequency domain, where the signal can be said to balance domain while spectral bandwidth is a measure of the spread or width of the frequency spectrum around the spectral centroid. They provide a measure of the amount the energy that is concentrated or dispersed across frequencies. Given that $X(f)$ represents the frequency domain of the signal obtained by implementing $x(t)$ into DFT in Formula (2), spectral centroid and spectral bandwidth can be formulated as follows:

$$SC = \frac{\sum_f f \cdot |X(f)|}{\sum_f |X(f)|}$$
$$SB = \sqrt{\frac{\sum_f (f - SC)^2 \cdot |f(X)|}{\sum_f |X(f)|}} \tag{12}$$

- Dominant frequency and total power: Dominant frequency is the frequency at which, in the Fourier spectrum of the signal, its magnitude is the largest, whereas total power is a notion of the overall energy, computed by summing over squared amplitudes. Given that $X(f)$ is the representation of time-domain signal $x(t)$, dominant frequency and total power can be calculated as follows:

$$f_d = \underset{f}{\operatorname{argmax}} |X(f)|$$
$$P_{total} = \sum_t |x(t)|^2 \tag{13}$$

## 3. Experiments

### 3.1. DeepFace for Gender and Age Prediction

In this study, it is necessary to utilize a sensor to automatically predict the physical properties. Therefore, DeepFace, a popular open-source framework for facial analysis, is utilized. DeepFace, which internally uses deep learning models pre-trained on very large datasets, offers a very high degree of accuracy for various facial attribute predictions, including age estimation and gender classification. Currently, the library supports several backend models, such as VGG-Face, Google FaceNet, and OpenFace. The VGG-Face backend provided by DeepFace was used in our implementation to analyze the first frame of every video regarding predicted age and gender due to its balance of speed and accuracy.

### 3.2. Experiment with Random Forest and Support Vector Regression for Multimodal Heart Rate Calculation

In this section, we outline the workflow utilized for our experiment, which encompasses a series of systematic steps designed to ensure a comprehensive evaluation of the machine learning models for heart rate estimation. The process begins with RoI extraction. Following this, feature engineering is employed to extract meaningful features from the signal data, including critical metrics such as the maximum, minimum, and peak-to-peak values; skewness; and kurtosis, which are vital for accurate model performance. All the

features collected from facial videos will be concatenated with physical features such as age, gender, weight, height, and BMI.

Next, we conduct model selection, opting for support vector regression (SVR) and random forest regression by tuning the hyperparameter. This tuning phase is executed using a grid search method, where various hyperparameters are evaluated to find the optimal configurations for each model. Hyperparameters for each model are given as follows:

- Hyperparameters for SVR: Kernel $\in$ {Linear, RBF}, $C \in$ {0.01, 0.1, 1, 10, 100}, $\gamma \in$ {0.01, 0.1, 1, 10, 100} and $\epsilon \in$ {0.001, 0.01, 0.1, 1, 10};
- Hyperparameters for Random Forest Regression: number of trees $n \in$ {50, 100, 200}, max depth $d \in \{none, 5, 10, 20\}$, and min samples per split $m \in$ {2, 5, 10, 20}.

All the processes are implemented using the scikit-learn library, which is a well-known and open-source tool for data analysis and machine learning implementation provided by Python.

Subsequently, we apply 5-fold cross-validation (CV), which is a statistical technique used to assess the performance of the tuned models on unseen data. This iterative process works by splitting the dataset into 5 folds, followed by training process using 4 folds as the training set and the remaining fold as the validation set. We repeat the training process until every fold becomes a training set. The result of the model can be obtained by averaging the results across all the folds [40]. Figure 6 below illustrates model evaluation using 5-fold CV.
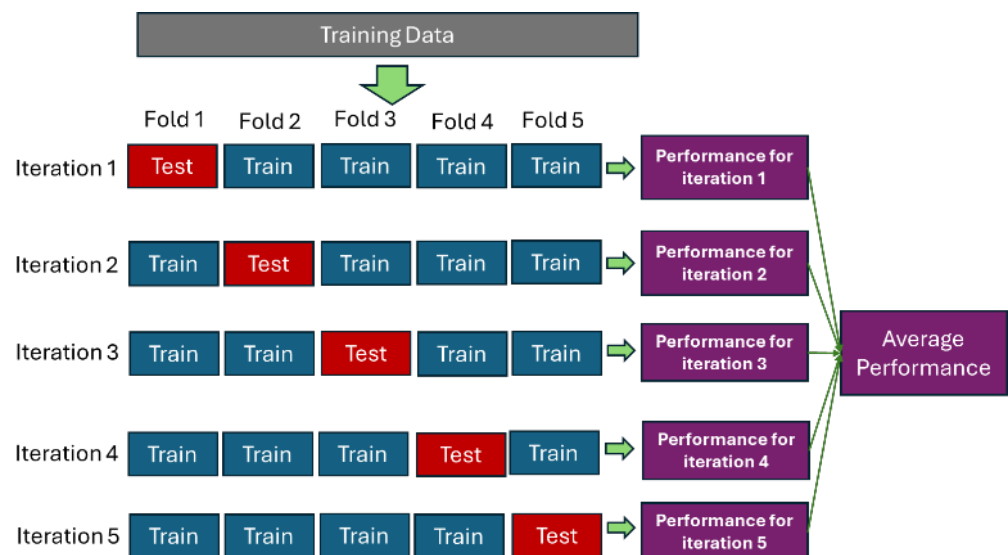


**Figure 6.** 5-fold cross-validation.

Finally, we evaluate the models using performance metrics such as mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) to quantify their predictive accuracy. Some commonly used techniques for measuring model performance in this study are MAE, RMSE, and MAPE, which can be formulated as follows:

$$MAE = \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{n} \tag{14}$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}} \tag{15}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i} \tag{16}$$

where $y_i$ is the actual value and $\hat{y}_i$ is the predicted value. For each metric, the lower the value, the better the performance of the model being tested. A detailed diagram illustrating this workflow is presented in Figure 7 to provide a visual representation of the steps taken throughout the experimental process.
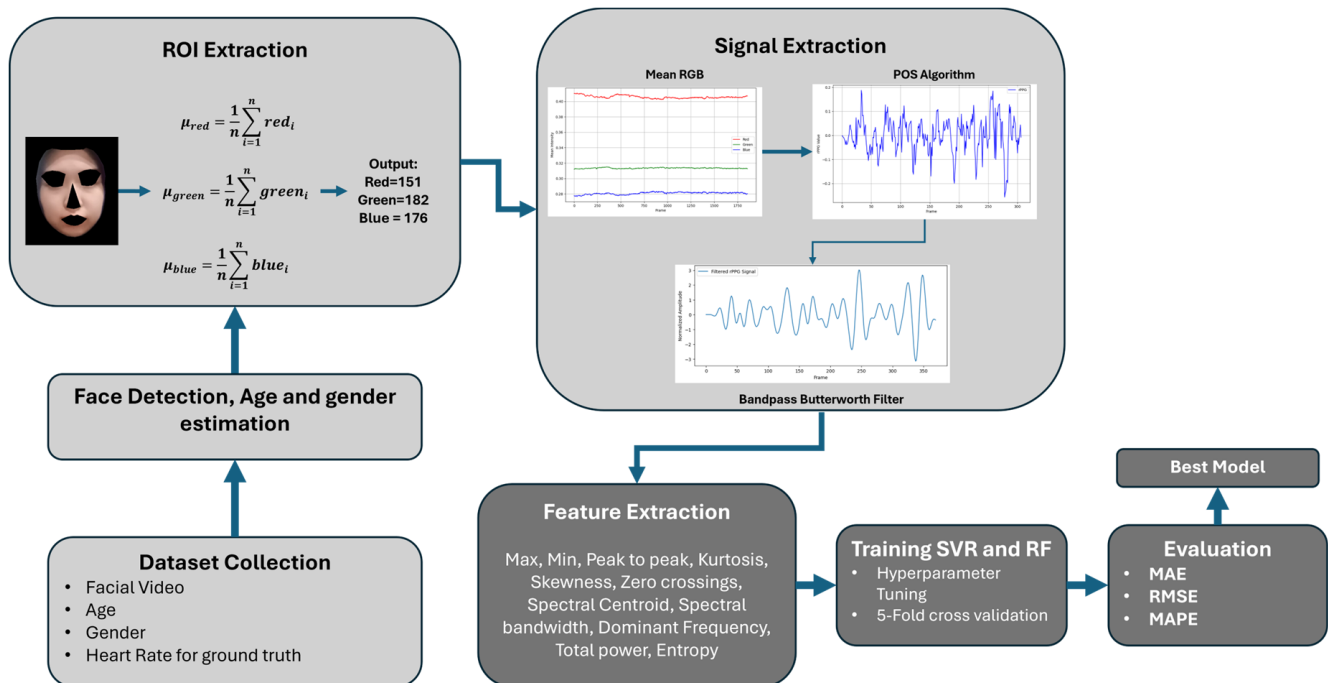


**Figure 7.** Workflow of the proposed method.

### 3.3. Performance Results

This study explored machine learning integration and multimodal approach for heart rate measurement by combining two types of data: signal features obtained from the facial video as listed in Equation (8) until Equation (13) were given the name "Video" and physical attributes, such as age, gender, height, and weight, were given the name "Physical". Two distinct machine learning models, namely support vector regression and random forest regression, are utilized. We systematically conducted ablation study by removing or altering specific video features and physical attributes to assess their impact on the MAE, RMSE, and MAPE. The focus was on understanding how each feature contributes to the model's overall accuracy in predicting the target variable. Table 1 shows the performance of all the ablation experiments conducted in this study. The columns Video and Physical use checkmarks (✓) and crosses (✗) to indicate whether the corresponding input type is included in each method. A checkmark means the input is included, while a cross means it is not. Generally, the random forest models perform better than the SVR with the best-performing method obtained when using the multimodal approach with the lowest MAE, RMSE, and MAPE of 3.057, 10.532, and 4.2%, respectively. In contrast, the best SVR is obtained when using physical information only with the MAE, RMSE, and MAPE of 3.251, 12.797, and 4.9%, respectively.

**Table 1.** Performance of proposed methods.

| Index | Method | Video | Physical | MAE | RMSE | MAPE | Hyperparameter Pairs |
|---|---|---|---|---|---|---|---|
| 1 | RF | ✓ | ✗ | 4.321 | 18.624 | 6% | $n = 70,\ d = 7,\ m = None$ |
| 2 | RF | ✗ | ✓ | 6.612 | 11.018 | 8.3% | $n = 120,\ d = 5,\ m = 2$ |
| 3 | RF | ✓ | ✓ | 3.057 | 10.532 | 4.2% | $n = 80,\ d = 7,\ m = 2$ |
| 4 | SVM | ✓ | ✗ | 7.601 | 18.624 | 10.2% | $C = 1000, \epsilon = 10, \gamma = 0.01,$ $kernel = linear$ |
| 5 | SVM | ✗ | ✓ | 3.251 | 12.797 | 4.9% | $C = 1000, \epsilon = 0.001, \gamma = 1,$ $kernel = RBF$ |
| 6 | SVM | ✓ | ✓ | 7.496 | 18.787 | 10.1% | $C = 1000, \epsilon = 10, \gamma = 0.01,$ $kernel = linear$ |

To gain more valuable insights from the random forest model, we checked predictive performance and interpretability using feature importances as shown in Figure 8 Feature importance provides a quick sense of how each feature is useful for the decisions of the model and, therefore, helps identify the most influential variables driving the predictions. This visualization is useful for making informed decisions about which features to retain, prioritize, or analyze further. By focusing on the most influential features, we can not only improve model performance but also uncover actionable insights for the problem domain.
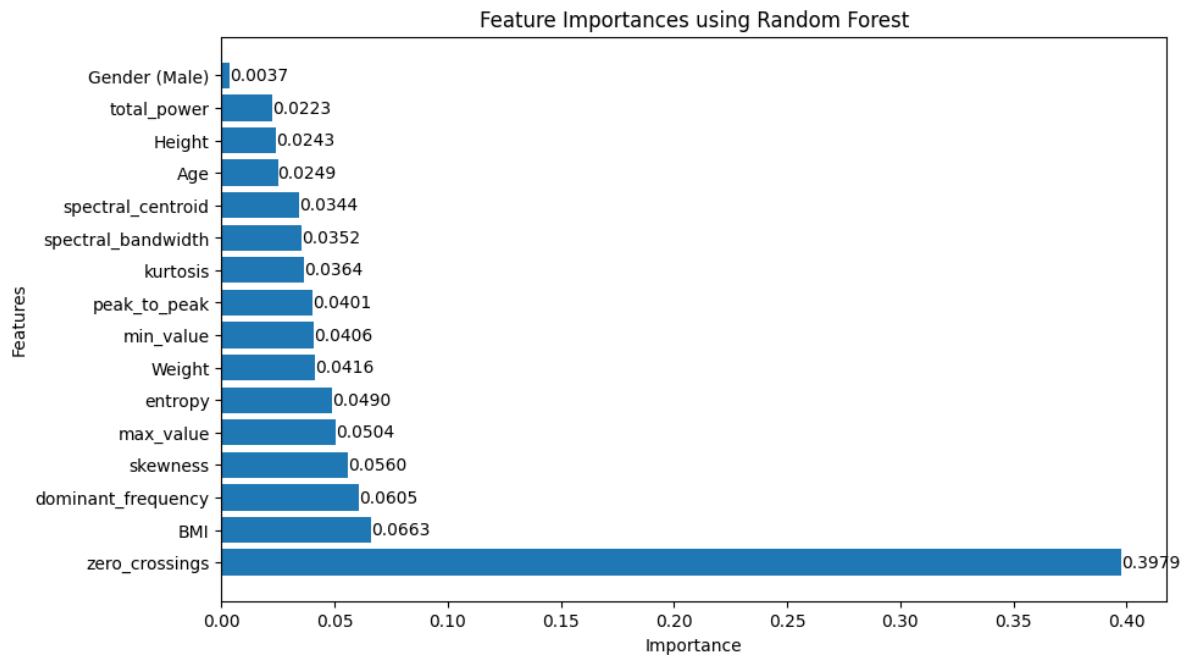


**Figure 8.** Feature importance.

In addition, Figure 9 shows Shapley additive explanation (SHAP) values, which completely explain how each feature contributes toward the output of the model for specific predictions. The values to the right, or more positive SHAP values, increase the prediction, while the values to the left, or negative SHAP values, decrease the prediction. Features that are further away from zero have a greater influence on a particular prediction. The SHAP values tell not only the direction or magnitude of each feature's effect but also consider the dependencies among the features, thus giving rich insight into the behavior of the model.
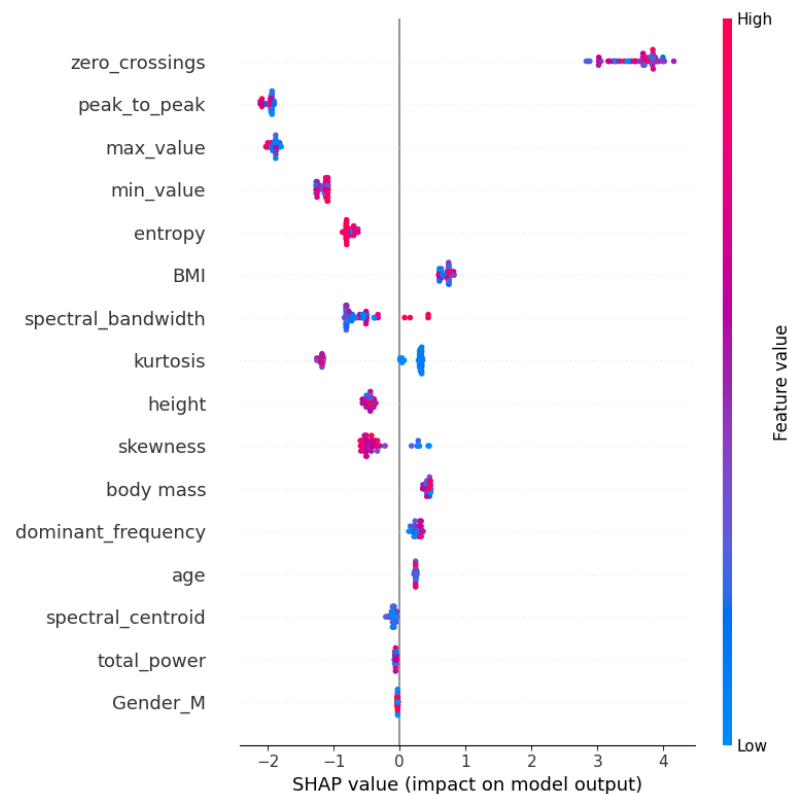
**Figure 9.** Shapley additive explanations.

For deeper analysis, Figure 10 shows a partial dependence plot (PDP), which provides insights into the effect of various features on the predictions of the random forest model on the multimodal data. Each subplot shows the partial dependence of a single feature, revealing how changes in its values impact the model's output. The *y*-axis, representing partial dependence, indicates the direction and strength of each feature's influence on the prediction. Higher values suggest a positive effect, while lower values suggest a negative one. Meanwhile, the *x*-axis shows the range of feature values, which are often normalized to facilitate comparison across different features.

For certain features like zero crossings, peak to peak, and BMI, the PDP curves display notable variations across specific ranges, suggesting that these features significantly impact the model's predictions in those areas. In contrast, features such as age, weight, height, skewness, and spectral bandwidth exhibit relatively flat lines, indicating that changes in these values have minimal effect on the prediction. The plot also shows the behavior of binary or categorical variables, like gender, where the flat line across 0 and 1 values suggests a limited impact on the outcome.

Features with sharp increases or decreases, like zero crossings, may indicate threshold effects, where a small change within a certain range can substantially affect predictions. This pattern suggests that the model may have learned non-linear relationships with these features, capturing complex interactions. Overall, this PDP grid highlights features such as zero crossings, peak to peak, and BMI which play a more prominent role in driving predictions, whereas others have less influence on the model's decision-making process.
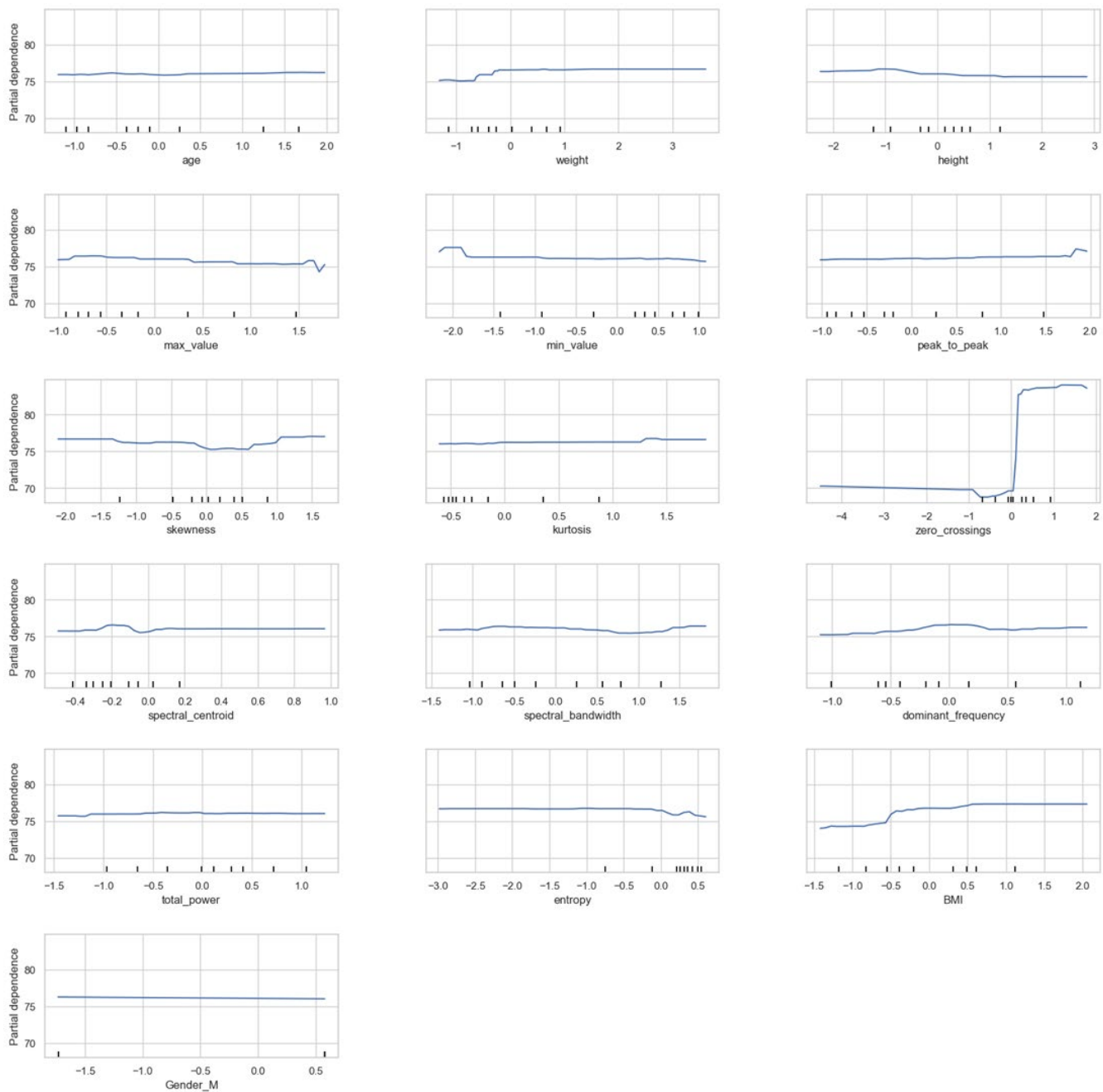
**Figure 10.** Partial dependence plot for random forest model.

*3.4. Comparison with State-of-the-Art Methods*

To evaluate the performance of the proposed method, we compared the result of our work to four different state-of-the-art methods explained in the introduction. The comparison is carried out by applying the state-of-the-art methods to the dataset used in this study and evaluated using the same metrics. It aims to compare how the proposed methodology measures the heart rate much more accurately than the state-of-the-art. It is easy to see from Figure 11 that our work considerably outperforms the existing state-of-the-art methods of remote photoplethysmography in heart rate measurement, evident from the lowest MAE, RMSE, and MAPE of 3.06, 10.53, and 4.02%, respectively. The random forest with multimodal approach outperforms the random forest without multimodal by 30% in terms of error, indicating the multimodal approach improves the remote heart rate measurement. Furthermore, the integration of random forest model and multimodal

approach outperforms the base POS with FFT by 72%, indicating that our approach is more reliable.
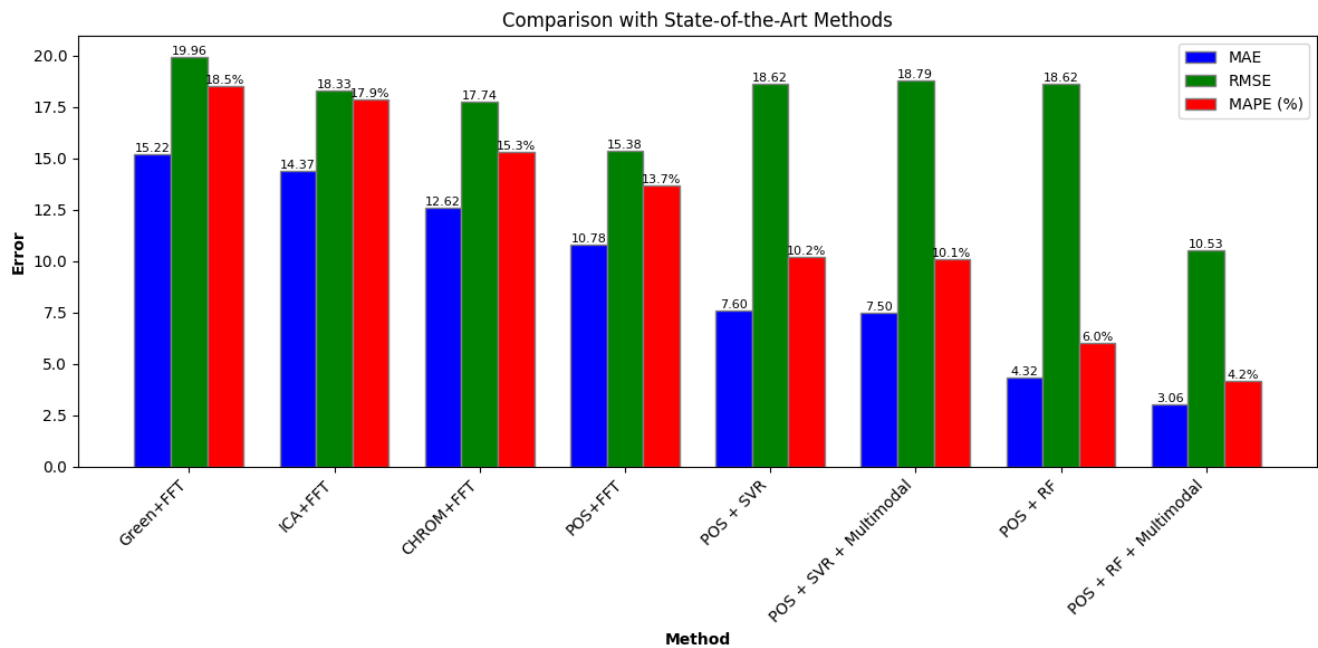


**Figure 11.** Comparison with state-of-the-art methods.

## 4. Discussion

In this study, we investigated the performance of machine learning integration and a multimodal approach for heart rate measurement by using the combination of the rPPG blood volume pulsatile signal extracted from the facial video and physical attributes such as age, gender, height, weight, and BMI. By combining these diverse data sources, we aim to improve heart rate estimation accuracy in non-contact scenarios, which is beneficial for telemedicine, fitness monitoring, and mass health screening.

This study explored machine learning integration and multimodal approach for heart rate measurement by combining two types of data: facial video as listed in Equation (8) until Equation (13) was given the name "Video" and physical attributes, such as age, gender, height, and weight, was given the name "Physical". Two distinct machine learning models, namely support vector regression and random forest regression, were utilized. We systematically conducted an ablation study by removing or altering specific video features and physical attributes to assess their impact on the MAE, RMSE, and MAPE. The focus was on understanding how each feature contributes to the model's overall accuracy in predicting the target variable. Figure 8 shows the performance of all the ablation experiments conducted in this study. Generally, the random forest models perform better than the SVR with the best-performing method obtained when using the multimodal approach with the lowest MAE, RMSE, and MAPE of 3.057, 10.532, and 4.2%, respectively. In contrast, the best SVR is obtained when using physical information only with the MAE, RMSE, and MAPE of 3.251, 12.797, and 4.9%, respectively.

Table 1 shows that the random forest regression with a multimodal approach is the best method which gives the lowest error with an MAE, RMSE, and MAPE of 3.057, 10.532, and 4.2%, respectively. This indicates the effectiveness of combining the two different data types in the prediction of heart rate. The models that did not include physical data or relied totally on video data had a higher error rate, underpinning the importance of physical attributes in enhancing the accuracy of the prediction.

In contrast, the SVR models perform comparably but are not better than the RF model that leverages both video and physical attributes. Interestingly, the SVR model with only physical data, Index 5, yielded a relatively low MAE of 3.251, while further addition of

either video or physical features degraded its performance. This suggests that while the SVR is successful with physical data, it is sensitive to the integration of video features, possibly due to the feature complexity or noise contributed by the multiple data types, which is the main drawback of SVR as mentioned by Farahmand, Desa, and Nilashi [41].

From these findings, it becomes clear that a multimodal approach, especially one that incorporates video and physical information, could greatly affect the performance of the model in estimating heart rate. The combined-data RF model's superior performance strengthens the promise of multimodal machine learning methods for improvement in the accuracy of remote heart rate measurement systems. This has led to the conclusion that multiple sources of data contribute to the overall better performance for applications, such as rPPG, when compared with single-modality methods.

These conclusions are further supported by the feature importance and SHAP value analyses of the best-performing RF scenario shown in Figure 9. First, it shows zero crossings as the most critical feature, with its importance score way higher than the other features, valued at 0.3979; hence, it contributed most to the model's predictions. Physical features such as BMI and dominant frequency also contribute to key roles but at a lower degree compared to the others. The SHAP value plot in Figure 9 offers more granularity on this, pointing out that zero crossings tend to have a consistent, high positive impact on the model output, while the impact of features such as BMI and peak to peak varies depending on their value. The domination of zero crossings as the most important feature in this study is due to its ability to measure the frequency of estimation, which is correlated with the heart rate.

Coupled with the discussion of feature importance, this suggests that the better performance of the RF model is indeed derived from its ability to exploit both video-derived features such as zero crossings and physical attributes like BMI. Allowing it to capture the complex interaction between these multimodal features tends to yield a low error rate in heart rate prediction. The SHAP analysis points out how these key features weigh in the output from the model, further reinforcing the fact that the conclusion must be about the inclusion of both video and physical data for the best performance. These results align with our hypothesis that physical attributes such as BMI influence the heart rate prediction [19–21].

The comparison in Figure 11 shows that the combination of random forest regression, POS algorithm, and multimodal data, which includes both video and physical attributes, achieves the lowest error rates among all the methods evaluated, with an MAE of 3.057, RMSE of 10.532, and MAPE of 4.2%. These results significantly outperform the other state-of-the-art methods, such as Green + FFT, ICA + FFT, and CHROM + FFT, which show considerably higher error rates. For instance, the Green + FFT method exhibits an MAE of 15.217 and a MAPE of 18.5%, highlighting the vast improvement achieved by integrating multimodal data with the random forest model.

## 5. Limitations and Future Works

Despite the promising results, there are some limitations in this study. First, the dataset used in this work is relatively small because no open-source site provides facial videos along with physical attributes information. Due to limited access to a large multimodal dataset, there is a risk that the model may be overfitting to the dataset used in this study. Future work should prioritize the collection of a larger, more diverse dataset to improve the model's generalizability and robustness across different populations.

Furthermore, an inaccurate physical attribute feed to the system such as the input of false weight or height may influence the result because measuring heart rate is based on false information. The limitation of this study is one that should be resolved in the future by collecting more multimodal datasets and aiming to develop a model that automatically predicts physical attributes to avoid manual inputting with the intent of avoiding feeding false information to the measurement system. Lastly, the rPPG signals extracted from the video are sensitive to lighting and environmental conditions, which could affect the accuracy

of heart rate estimation in non-controlled settings. Variations in light sources, shadows, and even skin tone may introduce noise in the video data. Future work could involve improving robustness to such factors by implementing data normalization techniques or training the model with data captured under diverse lighting conditions.

## 6. Conclusions

In this study, we aim to make noninvasive heart rate measurements more robust and reliable. We propose a machine learning integration into multimodal data including facial video, age, gender, weight, height, and BMI. The plane orthogonal to skin algorithm is utilized to extract blood volume pulsatile signal from the RGB facial video. Region of interest for video recording covers facial areas without eyes, nose, and lips. Signal filtering is implemented to denoise the obtained signal followed by signal extraction. Several features are selected, such as the maximum, minimum, and peak-to-peak value; skewness; kurtosis; zero crossings; spectral centroid; spectral bandwidth; dominant frequency; total power; and entropy value of the signal. Two machine learning models, support vector regression (SVR) and random forest regression are compared for this task. The experimental results demonstrate that incorporating a multimodal approach enhances model performance, with the random forest regression model achieving the best result surpassing the state-of-the-art methods. Furthermore, our method improves the robustness of non-contact heart rate measurement.

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

**Institutional Review Board Statement:** Ethical review and approval were waived for this study due to all participants providing informed consent and the assurance that their data would remain protected.

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

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author due to privacy and ethical restrictions, as the data involve human subjects.

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

## References

1. Yehezky, H.; Bustamam, A.; Hermawan, H. Enhancing ECG Signal Data through Denoising Features with Transformer Generative Adversarial Networks for Model Classication 1D-CNN. *Res. Sq.* **2023**. [CrossRef]
2. Atmadja, H.Y.N.; Bustamam, A. Hermawan Generated Tabular Data with Multi-GANs for Arrhythmia Classification Based on DNN Models. In Proceedings of the 2022 International Conference of Science and Information Technology in Smart Administration, ICSINTESA 2022, IEEE, Denpasar, Indonesia, 10–12 November 2022; pp. 69–74.
3. Shyam, A.; Ravichandran, V.; Preejith, S.P.; Joseph, J.; Sivaprakasam, M. PPGnet: Deep Network for Device Independent Heart Rate Estimationfrom Photoplethysmogram. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019.
4. Schiweck, C.; Piette, D.; Berckmans, D.; Claes, S.; Vrieze, E. Heart Rate and High Frequency Heart Rate Variability during Stress as Biomarker for Clinical Depression. A Systematic Review. *Psychol. Med.* **2019**, *49*, 200–211. [CrossRef] [PubMed]
5. Qiao, D.; Ayesha, A.H.; Zulkernine, F.; Jaffar, N.; Masroor, R. ReViSe: Remote Vital Signs Measurement Using Smartphone Camera. *IEEE Access* **2022**, *10*, 131656–131670. [CrossRef]
6. Hertzman, A.B. Observations on the Finger Volume Pulse Recorded Photoelectrically. *Am. J. Physiol.* **1937**, *119*, 334–335.

7. Soldatos, J. *Big Data and Artificial Intelligence in Digital Finance*; Kyriazis, D., Ed.; Springer International Publishing: Cham, Switzerland, 2022.

8. Upadhyay, B.R.; Baral, A.B.; Torlak, M. Vital Sign Detection via Angular and Range Measurements with MmWave MIMO Radars: Algorithms and Trials. *IEEE Access* **2022**, *10*, 106017–106032. [CrossRef]

9. Pnevmatikakis, A.; Kanavos, S.; Perikleous, A.; Kyriazakos, S. Risk Assessment for Personalized Health Insurance Products. In *Big Data and Artificial Intelligence in Digital Finance*; Springer: Cham, Switzerland, 2022; pp. 281–294. [CrossRef]

10. Verkruysse, W.; Svaasand, L.O.; Stuart Nelson, J. Remote Plethysmographic Imaging Using Ambient Light. *Opt. Express* **2008**, *16*, 2878–2886. [CrossRef]

11. Poh, M.-Z.; Mcduff, D.J.; Picard, R.W. Non-Contact, Automated Cardiac Pulse Measurements Using Video Imaging and Blind Source Separation. References and Links. *Opt. Express* **2010**, *18*, 10762–10774. [CrossRef]

12. De Haan, G.; Jeanne, V. Robust Pulse Rate from Chrominance-Based RPPG. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2878–2886. [CrossRef]

13. Wang, W.; Den Brinker, A.C.; Stuijk, S.; De Haan, G. Algorithmic Principles of Remote PPG. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 1479–1491. [CrossRef]

14. Huang, X.; Lv, J. Resilient Remote Heart Rate Measurement from Partially Masked Faces. In Proceedings of the International Joint Conference on Neural Networks, IEEE, Gold Coast, Australia, 18–23 June 2023.

15. Niu, X.; Zhao, X.; Han, H.; Das, A.; Dantcheva, A.; Shan, S.; Chen, X. Robust Remote Heart Rate Estimation from Face Utilizing-Temporal Attention. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019.

16. Kodithuwakkuge, J.; Sekhar, V.; Lamahewa, T.A. Blind Symbol Rate Estimation Using Convolutional Neural Networks for PSK Modulated Signals. In Proceedings of the 2021 15th International Conference on Signal Processing and Communication Systems, ICSPCS 2021—Proceedings, IEEE, Sydney, Australia, 13–15 December 2021.

17. Reiss, A.; Indlekofer, I.; Schmidt, P.; Van Laerhoven, K. Deep PPG: Large-Scale Heart Rate Estimation with Convolutional Neural Networks. *Sensors* **2019**, *19*, 3079. [CrossRef]

18. Kim, K.B.; Baek, H.J. Photoplethysmography in Wearable Devices: A Comprehensive Review of Technological Advances, Current Challenges, and Future Directions. *Electronics* **2023**, *12*, 2923. [CrossRef]

19. Quer, G.; Gouda, P.; Galarnyk, M.; Topol, E.J.; Steinhubl, S.R. Inter- and Intraindividual Variability in Daily Resting Heart Rate and Its Associations with Age, Sex, Sleep, BMI, and Time of Year: Retrospective, Longitudinal Cohort Study of 92,457 Adults. *PLoS ONE* **2020**, *15*, e0227709. [CrossRef]

20. Ehrenwald, M.; Wasserman, A.; Shenhar-Tsarfaty, S.; Zeltser, D.; Friedensohn, L.; Shapira, I.; Berliner, S.; Rogowski, O. Exercise Capacity and Body Mass Index—Important Predictors of Change in Resting Heart Rate. *BMC Cardiovasc. Disord.* **2019**, *19*, 307. [CrossRef]

21. Gonzales, T.I.; Jeon, J.Y.; Lindsay, T.; Westgate, K.; Perez-Pozuelo, I.; Hollidge, S.; Wijndaele, K.; Rennie, K.; Forouhi, N.; Griffin, S.; et al. Resting Heart Rate Is a Population-Level Biomarker of Cardiorespiratory Fitness: The Fenland Study. *PLoS ONE* **2023**, *18*, e0285272. [CrossRef] [PubMed]

22. Wu, T.; Schmitt, H.J.; Blazek, V. Photoplethysmography Imaging: A New Noninvasive and Noncontact Method for Mapping of the Dermal Perfusion Changes. *Opt. Tech. Instrum. Meas. Blood Compos. Struct. Dyn.* **2000**, *4163*, 62. [CrossRef]

23. Stricker, R.; Müller, S.; Gross, H.-M. Non-Contact Video-Based Pulse Rate Measurement on a Mobile Service Robot. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014.

24. De Haan, G.; Van Leest, A. Improved Motion Robustness of Remote-PPG by Using the Blood Volume Pulse Signature. *Physiol. Meas.* **2014**, *35*, 1913. [CrossRef] [PubMed]

25. Kim, J.K.; Mok Ahn, J. Design of an Optimal Digital IIR Filter for Heart Rate Variability by Photoplethysmogram. *Int. J. Eng. Res. Technol.* **2018**, *11*, 2009–2021.

26. Tarvainen, M.P.; Ranta-aho, P.O.; Karjalainen, P.A. An Advanced Detrending Method with Application to HRV Analysis. *IEEE Trans. Biomed. Eng.* **2002**, *49*, 172–175. [CrossRef]

27. Steiglitz, K. *Digital Signal Processing*; Courier Dover Publications: Mineola, NY, USA, 2020.

28. Guler, S.; Golparvar, A.; Ozturk, O.; Dogan, H.; Kaya Yapici, M. Optimal Digital Filter Selection for Remote Photoplethysmography (RPPG) Signal Conditioning. *Biomed. Phys. Eng. Express* **2023**, *9*, 027001. [CrossRef]

29. Burden, R.L. *Numerical Analysis*; Brooks/Cole: Pacific Grove, CA, USA, 2011.

30. Bustamam, A.; Sarwinda, D.; Abdillah, B.; Kaloka, T.Q. Detecting Lesion Characteristics of Diabetic Retinopathy Using Machine Learning and Computer Vision. *J. Electron. Syst.* **2020**, *10*, 23–33. [CrossRef]

31. Octaria, E.A.; Siswantining, T.; Bustamam, A.; Sarwinda, D. Kernel PCA and SVM-RFE Based Feature Selection for Classification of Dengue Microarray Dataset. *AIP Conf. Proc.* **2020**, *2264*, 030004.

32. Açıkkar, M.; Altunkol, Y. A Novel Hybrid PSO- and GS-Based Hyperparameter Optimization Algorithm for Support Vector Regression. *Neural Comput. Appl.* **2023**, *35*, 19961–19977. [CrossRef]

33. Bustamam, A.; Musti, M.I.S.; Hartomo, S.; Aprilia, S.; Tampubolon, P.P.; Lestari, D. Performance of Rotation Forest Ensemble Classifier and Feature Extractor in Predicting Protein Interactions Using Amino Acid Sequences. *BMC Genom.* **2019**, *20*, 950. [CrossRef]

34. Testas, A. Hyperparameter Tuning with Scikit-Learn and PySpark. In *Distributed Machine Learning with PySpark*; Apress: Berkeley, CA, USA, 2023; pp. 417–440. [CrossRef]

35. Kim, D.; Lee, K.; Sohn, C.B. Assessment of ROI Selection for Facial Video-Based RPPG. *Sensors* **2021**, *21*, 7923. [CrossRef]

36. Sarwinda, D.; Siswantining, T.; Bustamam, A. Classification of Diabetic Retinopathy Stages Using Histogram of Oriented Gradients and Shallow Learning. In Proceedings of the 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA), IEEE, Tangerang, Indonesia, 1–2 November 2018.

37. Oppenheim, A.V.; Willsky, A.S.; Nawab, S.H.; Ding, J.J. *Signals and Systems*; Prentice Hall: New York, NY, USA, 1997; Volume 2.

38. Singh, A.K.; Krishnan, S. ECG Signal Feature Extraction Trends in Methods and Applications. *BioMed. Eng. OnLine* **2023**, *22*, 22. [CrossRef] [PubMed]

39. Krishnan, S.; Athavale, Y. Trends in Biomedical Signal Feature Extraction. *Biomed. Signal Process. Control* **2018**, *43*, 41–63. [CrossRef]

40. Anguita, D.; Ridella, S.; Rivieccio, F. K-Fold Generalization Capability Assessment for Support Vector Classifiers. In Proceedings of the 2005 IEEE International Joint Conference on Neural Networks, Montreal, QC, Canada, 31 July–4 August 2005; Volume 2, pp. 855–858. [CrossRef]

41. Farahmand, M.; Desa, M.I.; Nilashi, M. A Combined Data Envelopment Analysis and Support Vector Regression for Efficiency Evaluation of Large Decision Making Units. *Int. J. Eng. Technol.* **2014**, *6*, 2310–2321.