

Article

Non-Invasive Classification of Blood Glucose Level Based on Photoplethysmography Using Time–Frequency Analysis

Ernia Susana ¹, Kalamullah Ramli ^{1,*}, Prima Dewi Purnamasari ¹ and Nursama Heru Aprianoro ²

¹ Department of Electronic Engineering, Universitas Indonesia, Depok 16424, Indonesia

² Department of Radiodiagnostic and Radiotherapy, Poltekkes Kemenkes Jakarta II, Jakarta 12120, Indonesia

* Correspondence: kalamullah.ramli@ui.ac.id; Tel.: +62-856-9009-843

Abstract: Diabetes monitoring systems are crucial for avoiding potentially significant medical expenses. At this time, the only commercially viable monitoring methods that exist are invasive ones. Since patients are uncomfortable while blood samples are being taken, these techniques have significant disadvantages. The drawbacks of invasive treatments might be overcome by a painless, inexpensive, non-invasive approach to blood glucose level (BGL) monitoring. Photoplethysmography (PPG) signals obtained from sensor leads placed on specific organ tissues are collected using photodiodes and nearby infrared LEDs. Cardiovascular disease can be detected via photoplethysmography. These characteristics can be used to directly affect BGL monitoring in diabetic patients if PPG signals are used. The Guilin People's Hospital's open database was used to produce the data collection. The dataset was gathered from 219 adult respondents spanning an age range from 21 to 86 of which 48 percent were male. There were 2100 sampling points total for each PPG data segment. The methodology of feature extraction from data may assist in increasing the effectiveness of classifier training and testing. PPG data information is modified in the frequency domain by the instantaneous frequency (IF) and spectral entropy (SE) moments using the time–frequency (TF) analysis. Three different forms of raw data were used as inputs, and we investigated the original PPG signal, the PPG signal with instantaneous frequency, and the PPG signal with spectral entropy. According to the results of the model testing, the PPG signal with spectral entropy generated the best outcomes. Compared to decision trees, subspace k-nearest neighbor, and k-nearest neighbor, our suggested approach with the super vector machine obtains a greater level of accuracy. The super vector machine, with 91.3% accuracy and a training duration of 9 s, was the best classifier.

Keywords: blood glucose level; machine learning; instantaneous frequency; spectral entropy; photoplethysmography



Citation: Susana, E.; Ramli, K.; Purnamasari, P.D.; Aprianoro, N.H. Non-Invasive Classification of Blood Glucose Level Based on Photoplethysmography Using Time–Frequency Analysis. *Information* **2023**, *14*, 145. <https://doi.org/10.3390/info14030145>

Academic Editor: Khalid Sayood

Received: 20 December 2022

Revised: 21 January 2023

Accepted: 2 February 2023

Published: 23 February 2023



Copyright: © 2023 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

Uncontrolled blood sugar levels in diabetics have the potential to seriously harm organs and result in consequences such as strokes that can lead to death. It is considered a “silent killer” because it can be fatal if not treated early [1,2]. The International Diabetes Federation predicted in 2017 that the prevalence of diabetes would increase internationally [3]. The two main types of diabetes, type 1 (T1D) and type 2 (T2D), have the same symptoms but differ in terms of causes and treatment. Unfortunately, many people with diabetes ignore it and do not realize how serious the side effects can be when their blood sugar levels increase. Type 2 diabetes is the most common and can be prevented or reduced by increasing awareness of its dangers and modifying to a healthy lifestyle [4]. Monitoring blood glucose levels regularly and administering diabetes treatments as needed can effectively prevent diabetes [5,6].

Currently, non-invasive monitoring with laboratory tests or a glucometer is the only method routinely performed and is believed to have a high degree of accuracy [7]. A

glucometer is commonly utilized in the home. Both procedures require using needles or finger pricks, which result in minimal tissue damage but discomfort to the patient.

A non-invasive procedure as an alternative to invasive testing is needed as a new method for monitoring BGL. At least four self-administered blood sugar checks are suggested daily, with an increased frequency if the patient's health is poor. However, some statistics indicate that between 40% and 50% of diabetics do not adhere to the recommendations [8,9]. Since this tool would be painless, comfortable, and relatively accurate and could be used repeatedly, it would increase patient compliance and encourage more discipline in monitoring their blood sugar.

Techniques arising from studies into the creation of non-invasive BGL monitoring or prediction systems are shown in Figure 1 [7,10,11]. In general, non-invasive techniques can be separated into two categories: sensors and optical. Sensor approaches detect glucose levels in biofluids such as saliva, tears, urine, and intracellular fluids [12]. The range of glucose concentrations in the intracellular fluid is comparable to that in the blood of healthy and diabetic patients. However, many studies [13] have shown a 4 to 10 min lag between intracellular fluid and blood glucose.

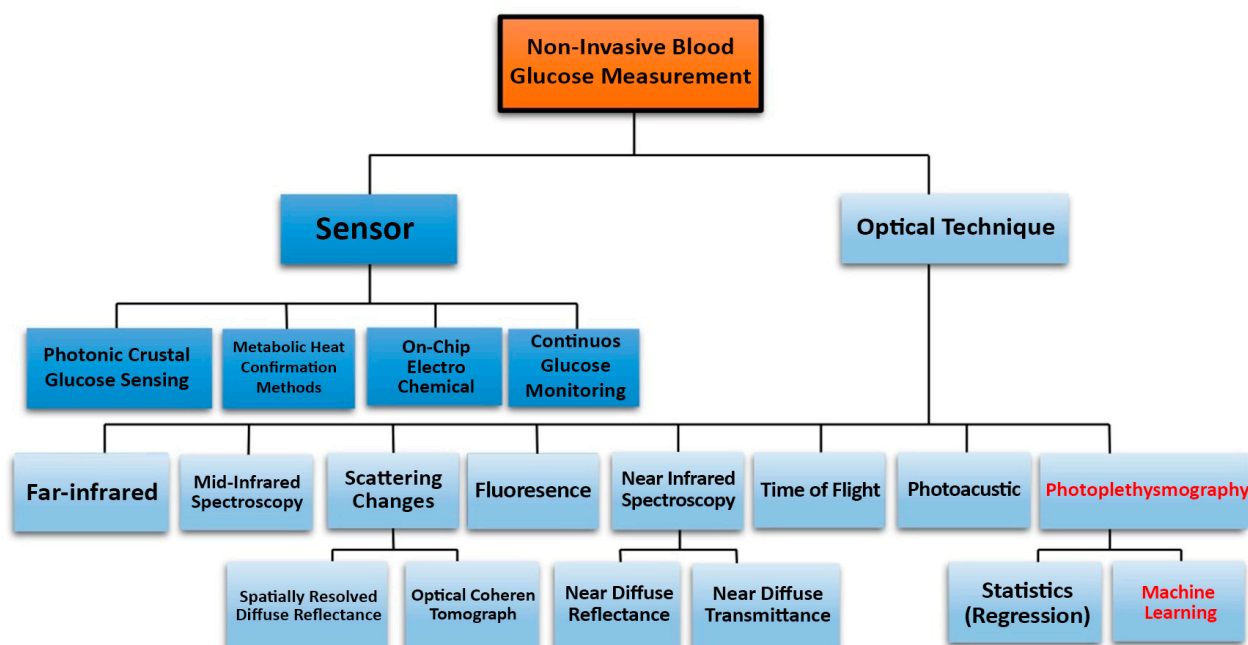


Figure 1. Non-invasive blood glucose measuring methods.

Optical approaches are evolving faster than sensors. As optical methods, near-infrared and mid-infrared spectroscopy, fluorescence, Raman, time of flight, and photoacoustic approaches have been developed in previous studies [10,12]. These non-invasive techniques are still undergoing study and development to meet market expectations [10]. Table 1 displays the strengths and weaknesses of each optical technique.

Photoplethysmography (PPG) is a new innovative non-invasive blood glucose monitoring technology that has been widely developed in recent years [14]. Optically based, PPG requires a light source to illuminate tissue and a photodetector to detect minute fluctuations in light intensity induced by perfusion-related capillary volume changes. PPG can operate in two transmission and reflection modes. From these signal intercepts, much information can be obtained, including pulse rate and oxygen levels in the blood.

Other important information that can be explored on the PPG signal [15] includes blood pressure [16], arterial stiffness [17], respiration [18], and diabetes [19]. Sensor placement is generally on the fingers, toes, and ears.

Table 1. Summary of optical technique technology [10,20,21].

Technology	Wavelength	Measurement Sites	Strengths	Weaknesses
Near Infrared Spectroscopy	750–2500 nm	Tongue, cheek, lip mucosa, forearm, ear lobe, and oral mucosa	Affordable, simple to implement	Humidity, pressure, temperature, the distribution of glucose impact accuracy, and other chemical substances interfere
Mid Infrared Spectroscopy	2500–10,000nm	Finger, skin, and oral mucosa	Very accurate, light, and low scattering	Inadequate skin penetration ability, and water absorption
Far Infrared Spectroscopy	30 μ m to 3 mm	Interstitial fluid (ISF)	Daily individual calibration is not required; the scattering is lower when compared to near infrared and mid infrared	It is not easy to differentiate between molecules other than water due to the strong absorption of water
Fluorescence	Ultraviolet light, visible light	Tears and human skin	High sensitivity and specificity to the presence of glucose and light scattering have no impact.	Sensitive to changes in oxygen and pH and prone to toxicity issues
Photoacoustic Spectroscopy	Ultraviolet light, NIR, and MIR	Aqueous humor, the forearm, and the finger	Unaffected by dispersed particles and resistant to water distortion	Low signal-to-noise ratio and affected by pulsation, acoustic noise, temperature fluctuations, and motion
Photoplethysmography	750–1500 nm	Ear lobe, toe, finger, and forehead	Simple, inexpensive sensor, and can be integrated with wearable devices and smartphone cameras	Unstable with movement and the characteristics of the resulting waves are affected by the conditions of blood circulation

In recent years, photoplethysmography (PPG) has gained popularity because of its ability as a simple and inexpensive non-invasive vital sign monitoring device. Another thing that supports PPG's popularity is the use of these sensors in wearable devices [20] and smartphones (webcams) [21,22] that has become possible due to recent technological advances. In addition, embedded system technology advancements and inexpensive sensor costs have enabled the production of PPG signal recorder prototypes.

The human body's tissues are located under the skin, the thickness of which varies depending on the location of the layers in the body. During systole, the volume of blood reaching the skin surface capillaries increases as the heart pumps blood throughout the body and lungs. This condition increases the absorption of light. The blood then returns to the heart via a network of veins, causing a decrease in capillary blood volume and light absorption.

Figure 2 illustrates the dual-component structure of the PPG signal, which consists of a pulsatile (AC = alternating current) and a non-pulsatile (DC = direct current) component. The AC component is vital because it contains all the information regarding the coordinated change in blood volume with each heartbeat. The AC component is superimposed on the "DC" component, which is the largest part of the PPG signal and has low-frequency fluctuations [23]. The DC component has information about the absorption of static elements

such as muscle, fat, tissue, and bone. The DC component provides vital data regarding respiration, venous flow, sympathetic nervous system activity, and thermoregulation [20]. Depending on body area and skin color, the AC component is often less than 0.2–2% of the DC component. The perfusion index (PI) is the ratio of pulsatile to non-pulsatile absorption [24].

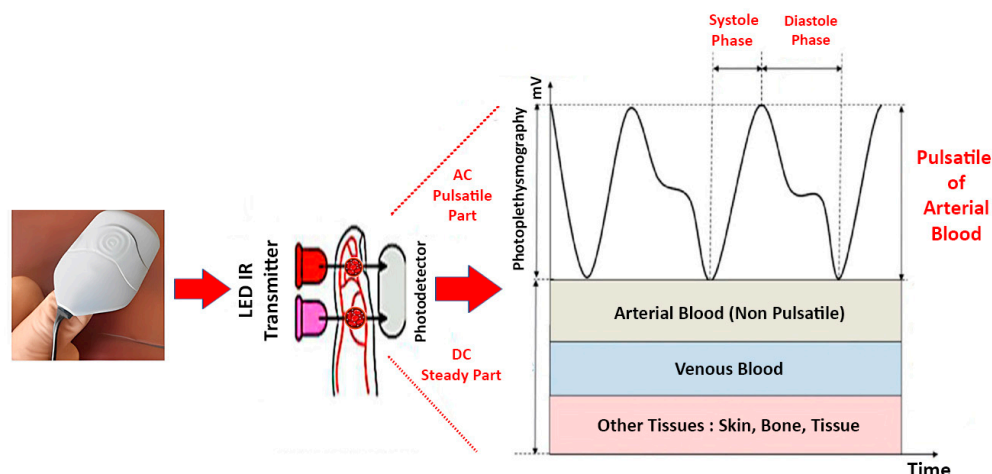


Figure 2. Dual component the PPG signal's structure.

The mechanism of the cardiac cycle alternates between systole (contraction) and diastole (relaxation) (diastolic). During the systolic phase, the heart muscle constricts to pump blood throughout the body's arteries as the heart beats. Contraction of the heart muscle will then put pressure on the arteries. After the ventricles have finished contracting, they will rest. During the relaxation phase (diastolic), the ventricles will fill with blood and become ready to carry out the following contraction process. The second waveform, which arises between the systolic and diastolic phases of the AC component, has a distinguishing and recognizable characteristic called the dicrotic notch [25].

Clinically, autonomic nervous system (ANS) dysfunction is a significant complication of diabetes mellitus. Changes in the autonomic nervous system will impact the heart rate variability (HRV) [26]. In each heartbeat, the heart rhythm varies. Heart rate variability refers to the regular change in milliseconds between heartbeats. HRV monitoring is used in many physiological contexts to evaluate cardiovascular autonomic function [27]. The autonomic nervous system (ANS), which controls various bodily functions and is made up of the sympathetic and parasympathetic nervous systems, can be better understood through heart rate variability (HRV) [28]. Unfortunately, acquisition of the ECG (electrocardiography) signals is required to obtain the R to R interval, requiring clinical supervision.

Theoretically, the PPG can determine the interval between heartbeats and heart rate variability [29]. Cardiac rhythm occurs when the mechanical action of the heart is linked with its electrical activity, although a time lag is necessary for transmitting a pulse wave. Research by Selvaraj et al. demonstrated a high correlation between the R to R ECG interval and the inter-beat interval (IBI). This study corroborates that HRV can be estimated reliably from the peak-to-peak (PP) interval-based method in both time and frequency domains. HRV calculations can be traced using PPG morphology [30]. Previous studies have found some association between HRV and blood sugar levels (BGL) [31].

There is much information about blood in the PPG signal. Many possibilities of data can be explored from the perspective of PPG morphology. The PPG pulse shape contains many intriguing characteristics. One of the features is the dicrotic notch between the systolic and diastolic phases. The effect of physiological changes on PPG waveforms is related to hemodynamics and glucose levels [32,33].

Diabetes can be identified using signal information as a classification feature. The output waves are then examined using the PPG technique to determine whether normal or

diabetic individuals have different wave characteristics. Figure 3 illustrates the differences in PPG morphology between healthy and diabetic participants. People with diabetes usually have PPG signal waves that tend to be bell-shaped without any subsequent peaks, whereas healthy subjects have a slightly elongated shape in the diastole phase [34]. When vascular resistance is unusually high because of atherosclerosis, diabetes, or any vascular pathology that narrows the vessels, the form of the PPG beat detected in the periphery (such as the fingertip) can change significantly from that in the middle artery. In that situation, there is a significant reduction in the blood flow velocity from the big arteries to the tiny capillaries. As a result, when the periphery is reached, the rising pulse wave of blood pressure spreads is delayed and may lose the secondary (dicrotic) peak [35].

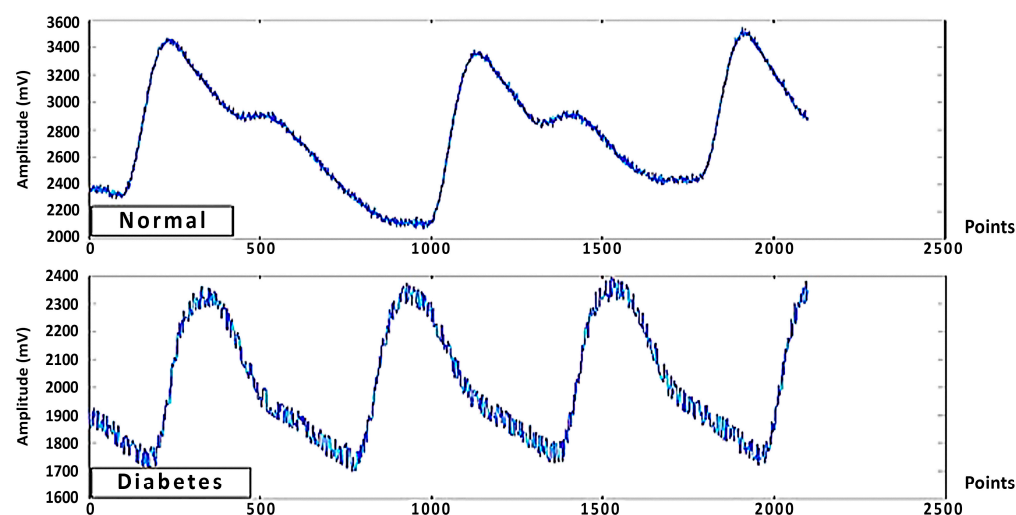


Figure 3. The PPG signal structure tends to differ between individuals with diabetes and healthy subjects.

Artificial intelligence (AI) is an intelligent system that can be utilized to mimic the ability of medical professionals to assist healthcare providers in their daily work and support decision making and problem solving. Predicting BGL is made possible by combining the PPG signal dataset with AI. Using PPG, machine learning [36,37] and deep learning [38,39] methods have similar effectiveness at predicting BGL classification. The lack of test samples is a frequent problem when implementing machine learning or deep learning.

The paper is structured as follows: Section 2 describes the materials and techniques, Section 3 explains the results, Section 4 analyzes the findings, and Section 5 provides a conclusion of the study.

2. Materials and Methods

As illustrated in Figure 4, we used an indirect technique by using patterns and sample points from PPG signals. The dataset is from the Guilin People's Hospital public database and the figshare public dataset, an online open access repository from the Guilin People's Hospital database [40,41]. PPG datasets from healthy and diabetic people are required to derive PPG signal patterns linked to BGL.

Using infrared light and transmission methods, a finger sensor is positioned on the tip of the finger, and PPG waves are utilized as the data. The information is then kept in a database in the text-based Numerical Notation (TXT) format. The corresponding application receives this real-time data through Bluetooth transmission. The PPG sensor model was the SEP9AF-2 (Korea, SMPLUS Company), which has a dual LED with wavelengths of 660 nm (Red light) and 905 nm (Infrared), a sampling rate of 1 kHz, a 12 bit analog digital converter (ADC), and a hardware filter design of 0.5 to 12 Hz bandpass. The probe's board has an integrated microprocessor with the model number MSP430FG4618 (Dallas, TX, USA, Texas Instruments business) [40].

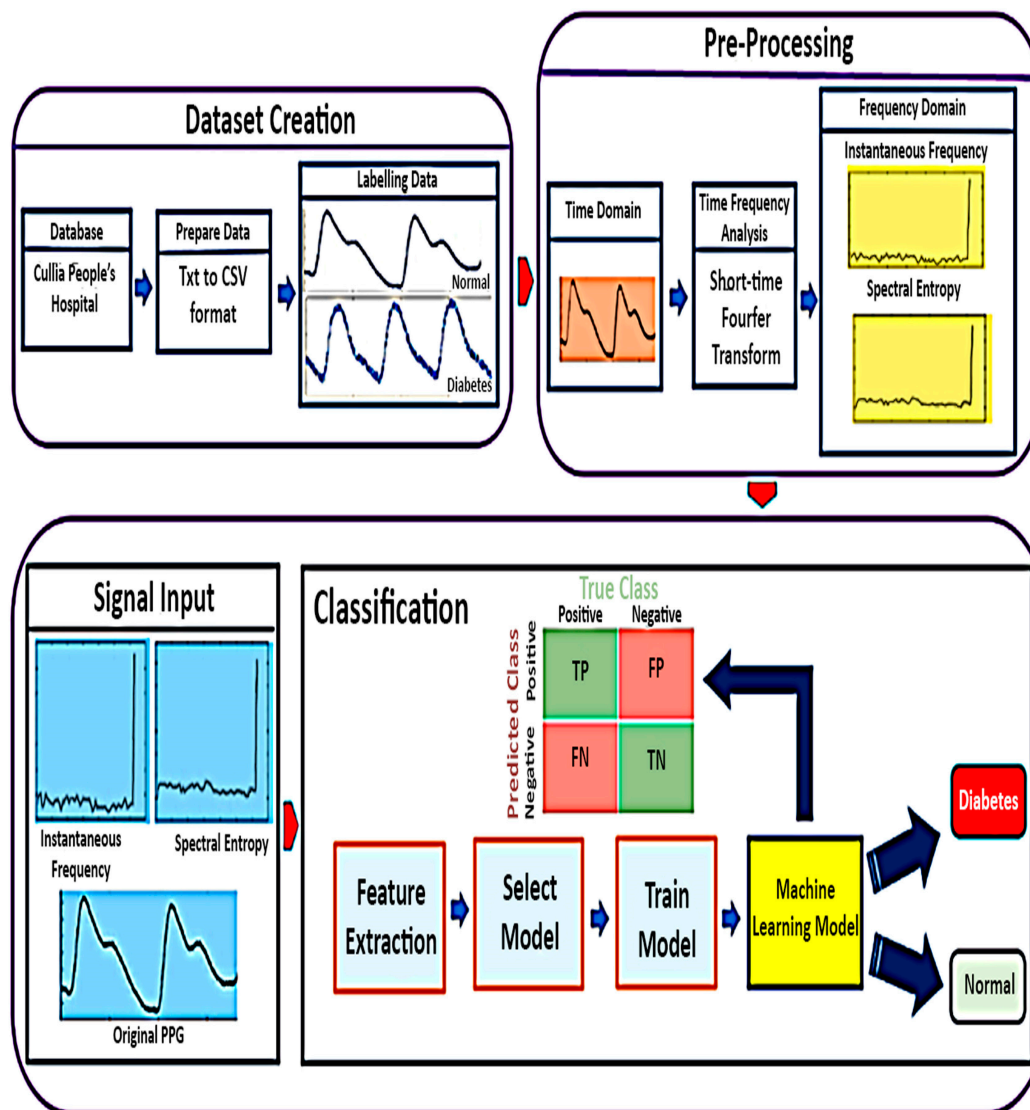


Figure 4. Block diagram representation of the suggested BGL classification approach. The three blocks represent the stages of processing: signal processing, classification results, and data collecting. There are 114 normal patients and 114 diabetics in each group's dataset. The instantaneous frequency and spectral entropy of each PPG waveform were obtained using temporal frequency analysis and a short-time Fourier transform (STFT). A machine learning classifier was then used to classify the BGL.

We used MATLAB to process the signals. Signal conversion requires a critical component called time–frequency analysis. The time–frequency analysis (TF) transforms information from the PPG signals into two TF moments in the time domain, i.e., instantaneous frequency (IF) and spectral entropy (SE), using a short-time Fourier transform (STFT). The time-dependent frequency of a signal is determined from the power spectrogram's initial instance using the IF function. Spectral entropy may provide exact information on the complexity of the signal based on the spectrum width. SE is used to compute the data that makes up the different frequency components.

2.1. Data Collection

The Guilin People's Hospital's open database was used to produce the data collection. This is available for download via the figshare repository [40,41]. This study was approved by the ethics committee of Guilin People's Hospital and Guilin University of Electronic Technology in China. All participants provided written and informed consent prior to

the study. Guilin People's Hospital data include 219 subjects covering the age range of 20–89 years. Information records include ID, gender, age, height, weight, systolic pressure, diastolic pressure, heart rate, BMI and common cardiovascular diseases (CVD) such as hypertension, diabetes, cerebral infarction and cerebrovascular disease. During the acquisition of the signal, the sampling waveform was placed at 1 kHz with a precision of 12 bits for the analog-to-digital conversion. There are 2100 sample points in each section of PPG data [40].

The experiment lasts for around 15 min in its entirety. Photoplethysmography (PPG) data gathering takes around 3 min. As seen in Figure 5, each segment has 2100 sample points, which each take 2.1 s to acquire [40].

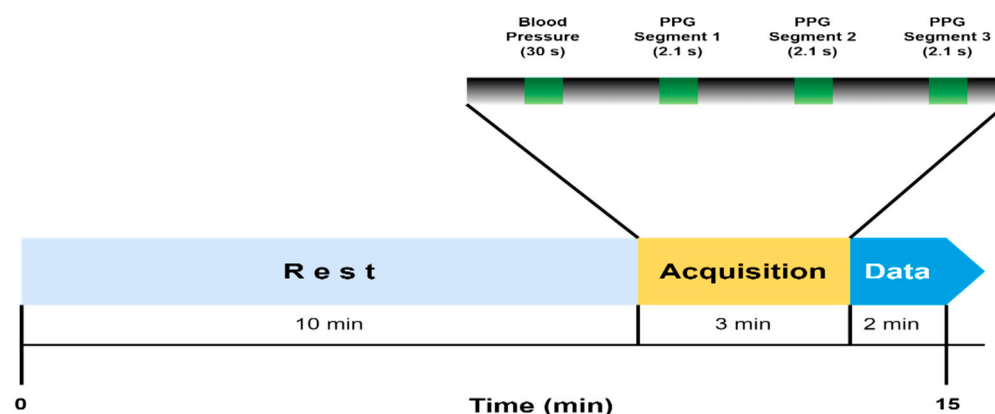


Figure 5. The whole experiment took about 15 min to complete. Each PPG segment lasted 2.1 s, and three were collected per individual.

We categorized the data into “signal” and “label” categories. As shown in Figure 3, there are two label values: “normal” and “diabetes”, whereas the signal includes an arrangement of cells containing a variety of PPG signals. The data are analyzed using the PPG signal quality throughout the data-collecting phase before being stored. We chose the signals using the waveform to predict poor PPG signals. Figure 6 shows the method for classification of PPG waveforms that was used to evaluate whether each PPG signal segment was acceptable and unacceptable.

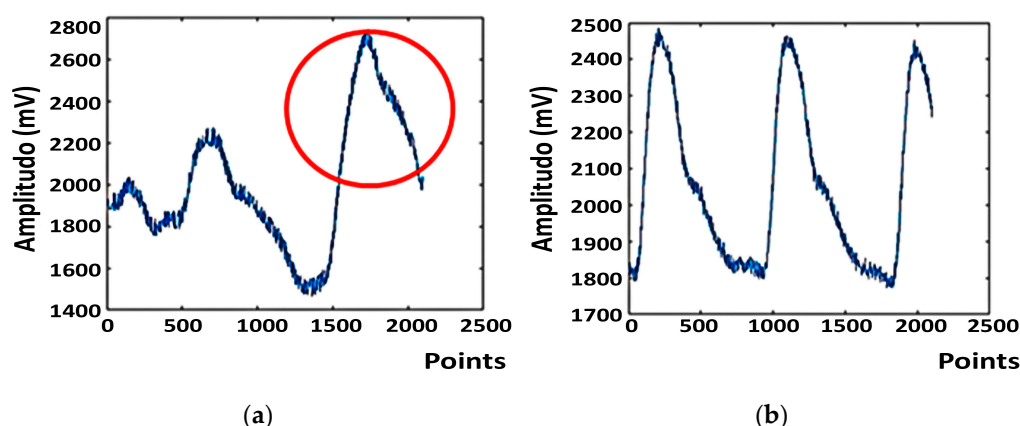


Figure 6. (a) Characteristics of unacceptable PPG signal. (b) Characteristics of acceptable PPG signal.

Perfect PPG waves have a diastolic notch and a distinct rhythm. The unacceptable category included several loud beats without a diastolic notch, while the acceptable category featured a clean beat without one. This strategy was inspired by the skewness signal quality index (SSQI), which differentiates the degree of asymmetries in the distribution around the mean [42]. Therefore, the signals are split into training to educate the classifier and a

testing set to evaluate the classifier's accuracy. Before including the datasets, the signal data from each classification level were balanced to ensure that each group contained the same training datasets (114 normal subjects and 114 diabetic subjects). The classifier's accuracy was assessed using 20% of the dataset, while the remaining 80% was used as a training set.

2.2. Time Frequency Analysis

The methodology of feature extraction from data may assist in increasing the effectiveness of classifier training and testing. The raw PPG data must be converted into a spectrogram before being used to build a feature representation. TF moments may be used to extract information from spectrograms. A PPG signal's temporal domain features may vary by very little, but the human eye is not sensitive enough to see them. Thus, methods for obtaining features in the time or frequency domain are necessary [43,44]. The signal is separated into small time windows for time–frequency analysis, resulting in a spectrum represented as a sliding window. Each pixel's frequency and time data serve as estimates for the PPG signal's intensity. The high frequency of the sound is easier to locate using a spectrogram. The PPG signal represents a complex interplay between cardiac activity, vascular relaxation, and microcirculation system status. In the temporal frequency domain, PPG signals thus include a lot of different information. PPG data information is modified in the frequency domain by the instantaneous frequency (IF) and spectral entropy (SE) moments using the time–frequency analysis.

The signal's instantaneous frequency (IF) is a property that often has significant practical importance. Physically, it only applies to monochromatic signals with a single frequency or a narrow range of frequencies that fluctuate over time [43]. This situation encourages the development of the IF concept that was recently proposed by Nobel Prize winner Dennis Gabor. The IF, as seen in Figure 7, determines the magnitude of the instantaneous phase shift in time or as a phase trace created from the instantaneous phase's first derivative. The IF is a brief sinusoidal frequency window that precisely matches the seismic trace. The importance of IF arises from the need to often examine signals whose spectral characteristics fluctuate over time, particularly the spectral peak frequencies [43]. The following is a definition of the instantaneous frequency for a real-time signal.

$$FI(t) = \frac{1}{2\pi} \frac{d}{dt} \phi(t) \quad (1)$$

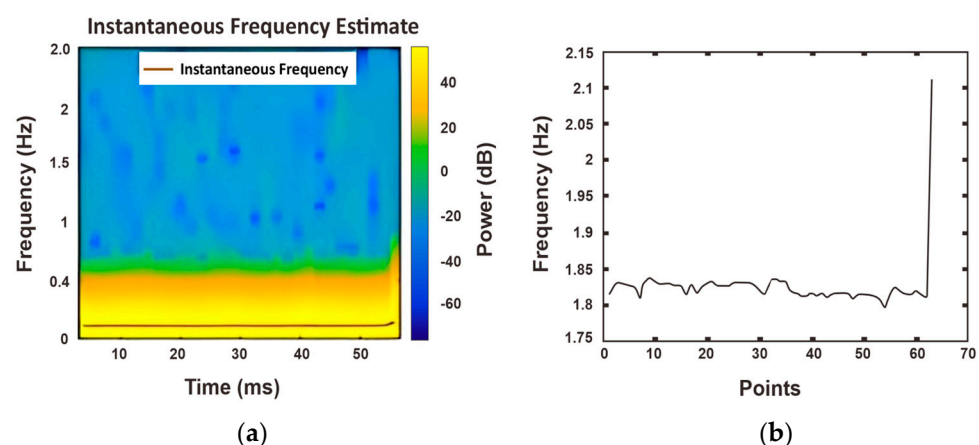


Figure 7. (a) The instantaneous frequency function estimates the time–dependent frequency of a signal as the first moment of the power spectrogram. (b) The instantaneous frequency from the spectrogram using short–time Fourier transforms over time windows. In this example, the function uses 63 time windows.

The spectral entropy function calculates the spectral entropy based on a power spectrogram. Spectrum entropy presents complete information on the signal structure based

on spectrum width. Wider spectrums, such as white noise, are linked to high entropy. Low entropy is reflected by a narrow spectrum, similar to a sum of sinusoids. In actuality, spectral entropy computes the specific functionality in each frequency component. The spectral entropy is calculated using the following equation.

$$x_i = \frac{X_i}{\sum_{i=1}^N X_i} \quad (2)$$

where x_i is the energy of the i th frequency component of the spectrum, $x = (x_1 \dots \dots, x_N)$ is the spectrum's probability mass function (PMF) and number of spectrum points.

In this work, the spectrogram is computed using 63 time frames using spectral entropy. Each time window's entropy is calculated from x using:

$$H = -\sum_{i=1}^M x_i \log_2 x_i \quad (3)$$

2.3. Classifier

The platform we used to choose the most effective machine learning classifier was Classification Learner from MATLAB. The platform provides an appropriate degree of accuracy for the algorithm's BGL categorization based on PPG data. The same dataset, as seen in Figure 8, was used to test each model. Several evaluation indices, including the F1 score, accuracy (Ac), recall (Re), specificity (Sp), precision (Pr), true negative (TN), false negative (FN), true positive (TP), false positive (FP), true negative (TN), and false positive (FN) were utilized to assess the trained models in their entirety. When the cost of a false positive (FN) is large, precision is a reasonable benchmark to use. Recall is beneficial when the expense of false negatives is significant. The F1 score is a comprehensive evaluation of the accuracy of a model, consisting of precision and recall. In comparison to the other variables, the normotension had the best F1 score overall.

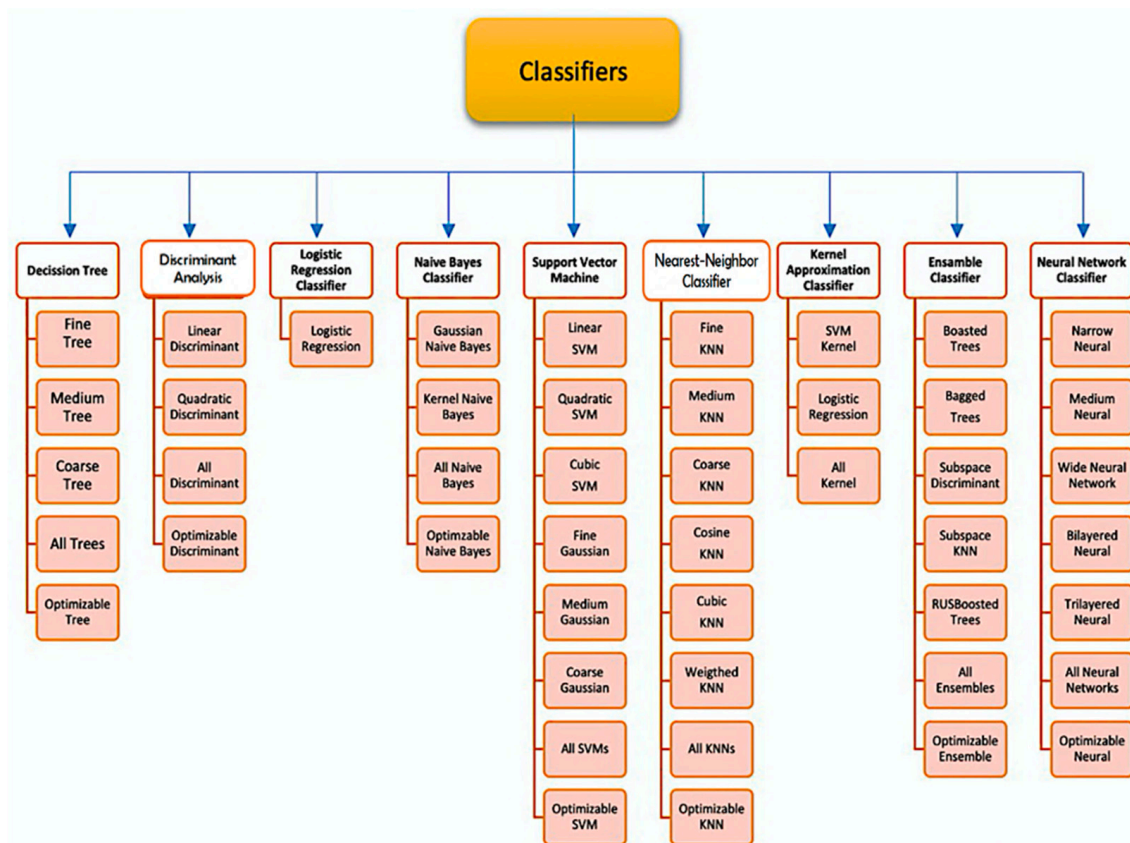


Figure 8. Classifier algorithms were tested.

2.4. Confusion Matrix

The amounts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) from the confusion matrix are used to calculate the five formulae mentioned above. The confusion matrix is a $N \times N$ matrix used to assess the effectiveness of a classification model, where N is the total number of target classes. In the matrix, the actual goal values are contrasted with those that the machine learning model anticipated. A confusion matrix is a table that lists how many guesses a classifier has correctly and incorrectly made. It is used to evaluate a classification model's effectiveness. The fundamental terminology in the confusion matrix is as follows:

- A. True negatives (TN) occur when the actual value and the prediction are both negative, and true positives (TP) occur when the actual value and the prediction are both positive.
- B. False positives (FP) occur when a positive outcome is predicted but the actual result is negative, also known as a Type 1 error.
- C. False negatives (FN) are when a negative outcome is predicted but the actual result is positive, sometimes referred to as a type 2 error.

3. Results

A training set (182 data) and a testing set (46 data) were created from the dataset for this investigation. Our proposed approach changes from deep learning to machine learning to shorten training time. Since deep learning is computationally intensive and takes some time to train, it might be difficult to evaluate and repeat experiments often in order to improve results. We evaluated the classification accuracy of three distinct data sources. The second and third data inputs are taken from the TF findings, which result in IF and SE with a frequency domain. The first data input is a pure PPG signal with a time domain that has not been treated. The three BGL-adjusted inputs shown in Figure 9 are contrasted in the following.

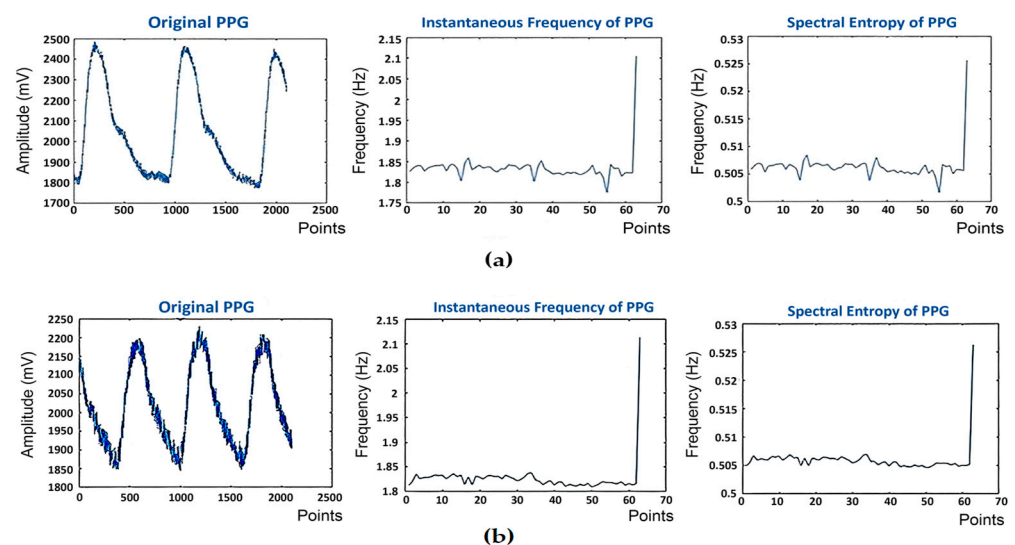


Figure 9. Comparison of the input signal's structure, original PPG, instantaneous frequency (IF), and spectral entropy (SE) in relation to BGL: (a) diabetics; (b) normal.

3.1. Original PPG

In the first experiment, we employed the original PPG signals to classify BGL. Each PPG segment comprises 2100 data points. *Data distribution* is a statistic that displays all data values and the frequency with which those occur to ensure no data outliers, as shown in Figure 10. We ran a comparative analysis of twenty-five classification algorithms to examine the training performance of this experiment. We present the three best machine

learning classifiers from the training phase in Table 2 and the ROC curve for the weighted KNN classifier during training in Figure 11.

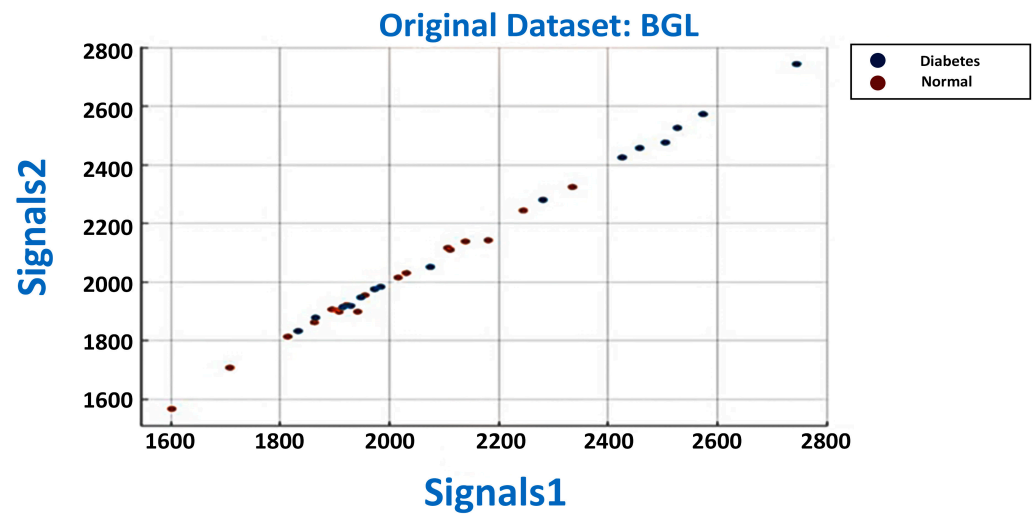


Figure 10. Data distribution of multiple measurements from PPG, which are divided into healthy objects and diabetic objects.

Table 2. Comparative analysis of training results.

Classification Model	Accuracy (%)	Estimation Speed (Observations/s)	Exercise Time (s)
Weighted KNN	85.7	~29	110.41
Super Vector Machine	85.2	~29	53.18
Linear Discriminant	84.1	~32	21.53

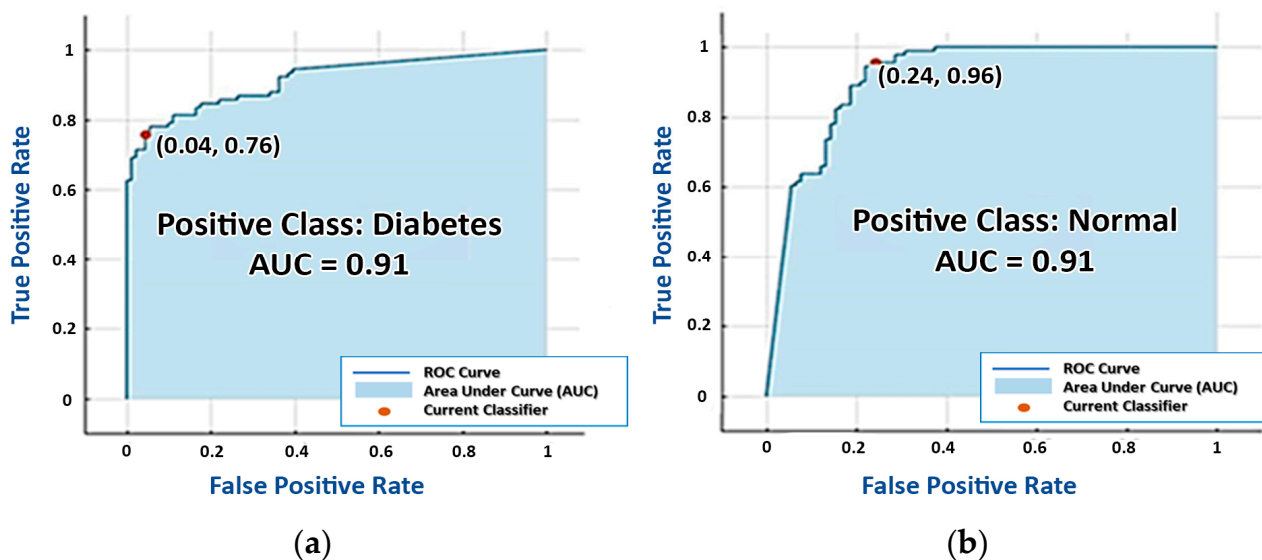


Figure 11. ROC curve for the weighted KNN classifier during training: (a) positive-class diabetes (b) positive-class normal.

We compared the testing performance based on the accuracy values. The weighted KNN classifier had an accuracy value of 86.9%, which was better than other classification methods. A confusion matrix was utilized to illustrate the performance of classifiers on datasets where the true value is known, as shown in Figure 12.

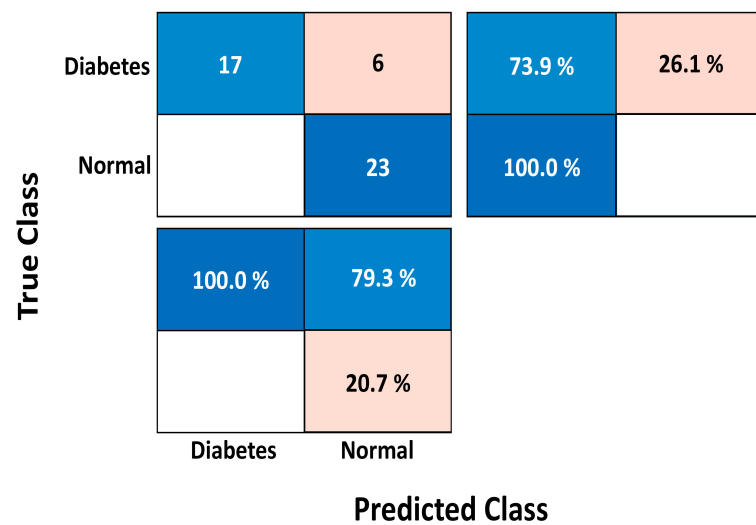


Figure 12. Confusion matrix for weighted KNN with testing accuracy: 86.96%.

Several classifier evaluation performance indicators were used, such as accuracy (Ac), recall (Re), specificity (Sp), precision (Pr), sensitivity (Se), and F1 score. The classification performance of our proposed method is shown in Table 3.

Table 3. Classification performance of weighted KNN classifier.

Trial	TP (%)	FP (%)	TN (%)	FN (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Recall (%)	Precision (%)	F1 Score (%)
Normal	100	0	73.9	26.1	86.96	79.31	100.00	79.31	100.00	88.46
Diabetes	73.9	26.1	100	0	86.96	100.00	79.31	100.00	73.90	84.99

3.2. Instantaneous Frequency

In the second experiment, we used instantaneous frequency PPG signals to classify BGL into categories. Each PPG segment consists of 63 sampling points. Data distribution is a statistic that displays all data values and the frequency in which these occur to verify that there are no data outliers, as shown in Figure 13.

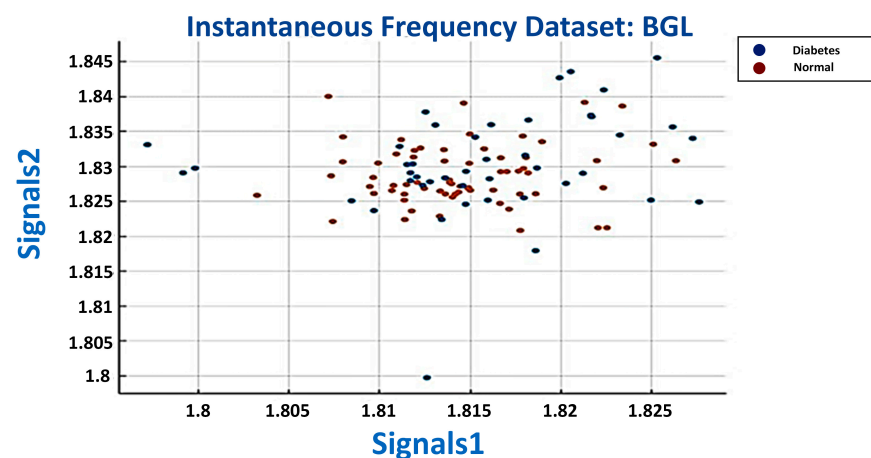
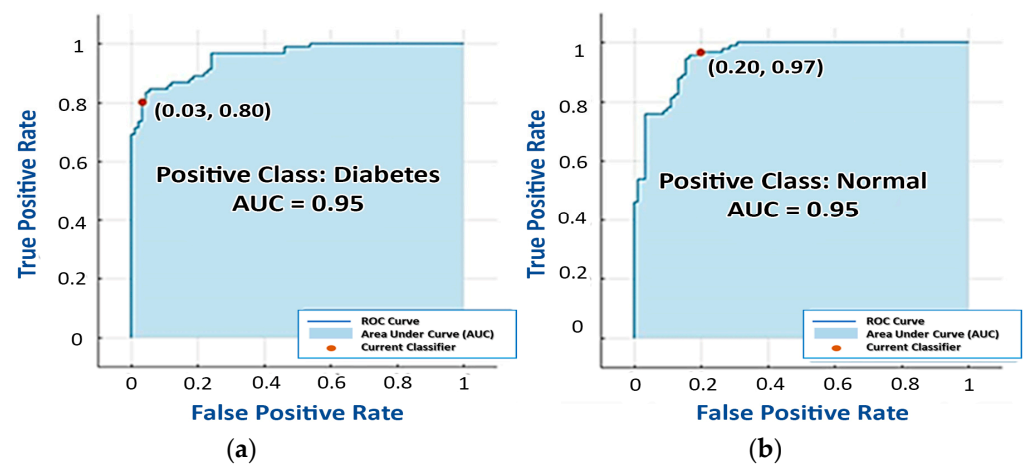


Figure 13. Data distribution of instantaneous frequency PPG signal.

We conducted a comparative analysis with twenty-five machine learning classifiers to assess the training performance of this work. We present the three best machine learning classifiers in the training phase in Table 4 and the ROC curve for the super vector machine classifier during training in Figure 14.

Table 4. Comparative analysis of training results.

Classification Model	Accuracy (%)	Estimation Speed (Observations/s)	Exercise Time (s)
Super Vector Machine	89.0	~930	1.67
Naive Bayes	86.8	~100	26.94
Ensemble Subspace KKN	82.4	~55	26.66

**Figure 14.** (a) Positive-class diabetes ROC curve; (b) positive-class normal ROC curve.

We evaluated testing outcomes on the basis of accuracy values. The super vector machine classifier has an accuracy value of 89.10%, which is better than other classification methods. As shown in Figure 15, a confusion matrix is employed to analyze the performance of classifiers on datasets for which the real value is known.

True Class	Diabetes	18	5	78.3 %	21.7 %
			23	100.0 %	
	Normal	100.0 %	82.1 %		
			17.9 %		
		Diabetes	Normal		
		Predicted Class			

Figure 15. Confusion matrix for super vector machine with testing accuracy: 89.1034%.

Several classifier evaluation performance indicators were used, such as accuracy (Ac), recall (Re), specificity (Sp), precision (Pr), sensitivity (Se), and F1 score. The classification performance of our proposed method is shown in Table 5.

Table 5. Classification performance of super vector machine classifier.

Trial	TP (%)	FP (%)	TN (%)	FN (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Recall (%)	Precision (%)	F1 Score (%)
Normal	100	0	78.3	21.7	89.15	82.17	100.00	82.17	100.00	90.21
Diabetes	78.3	21.7	100	0	89.15	100.00	82.17	100.00	78.30	89.33

3.3. Spectral Entropy

In the third experiment, we used spectral entropy PPG signals to classify BGL into categories. Each PPG segment consists of 63 sampling points.

Data distribution is a function that shows all the values of data and how often these values occur to ensure that there are no data outliers, as shown in Figure 16. We ran a comparative analysis with twenty-five machine learning classifiers to analyze the performance of the model of this study. We present the three best machine learning classifiers in the training phase in Table 6 and the ROC curve for the weighted KNN classifier during training in Figure 17.

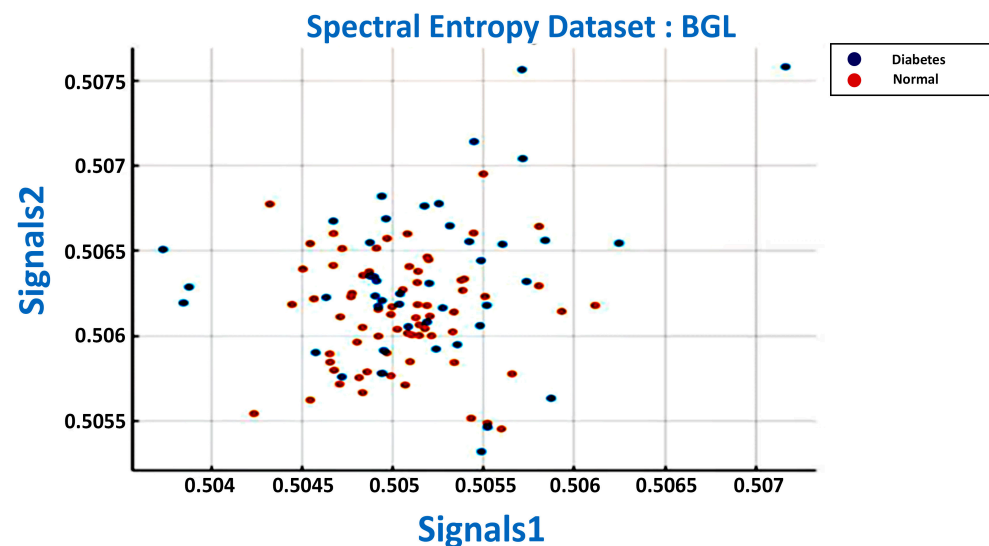
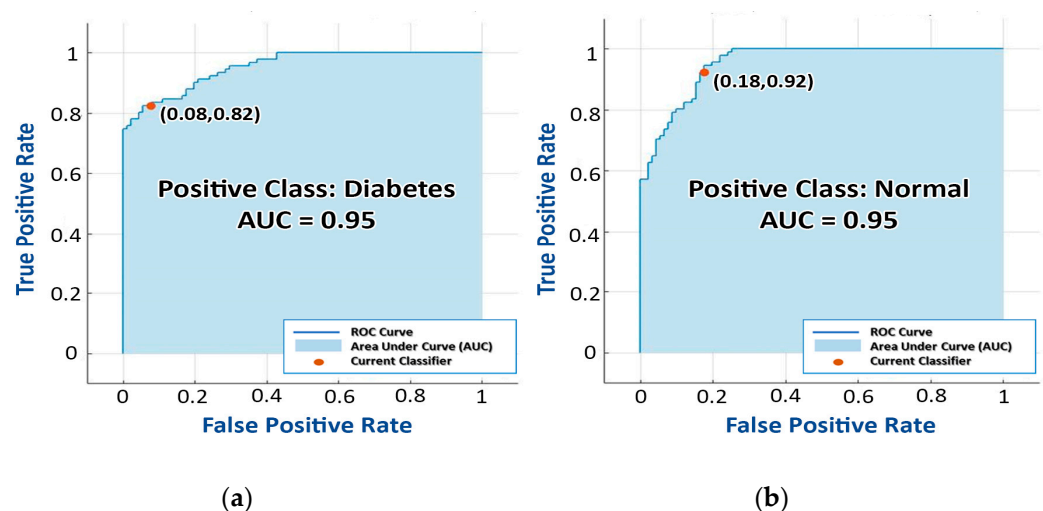
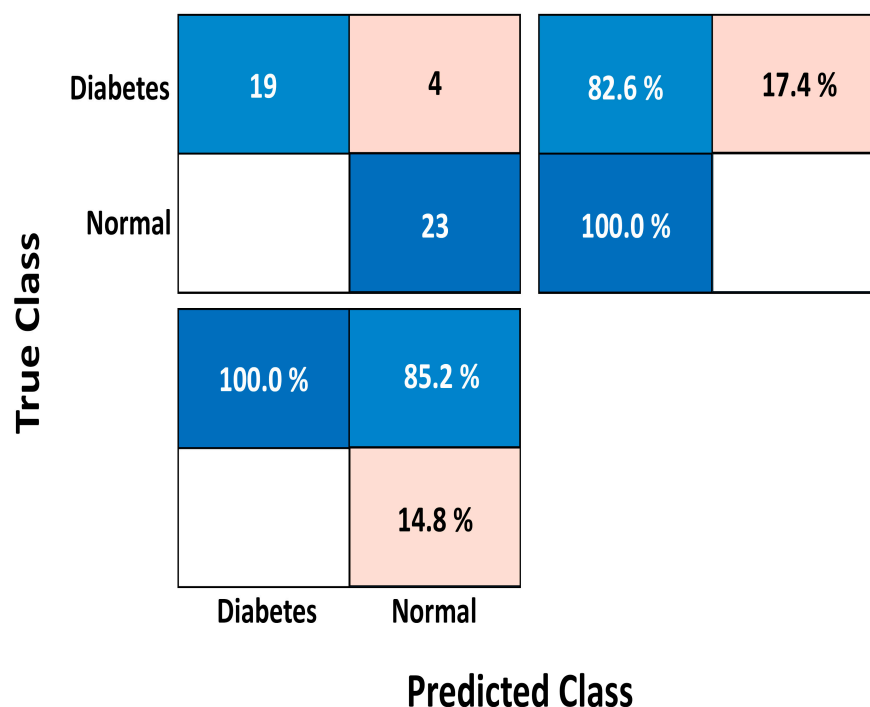
**Figure 16.** Data distribution of spectral entropy PPG signal.**Figure 17.** ROC curve for super vector machine classifier during training: (a) positive-class diabetes; (b) positive-class normal.

Table 6. Comparative analysis of training results.

Classification Model	Accuracy (%)	Estimation Speed (Observations/s)	Exercise Time (s)
Super Vector Machine	91.3	~520	9.25
Naive Bayes	86.8	~87	31.22
KKN	85.2	~400	9.98

We compared the results of testing based on the estimations of accuracy. The accuracy of the super vector machine classifier is 91.30%, which is greater than other classification models. A confusion matrix is applied to evaluate the performance of classifiers on datasets where the true value is known, as shown in Figure 18.

**Figure 18.** Confusion matrix for super vector machine with testing accuracy: 91.30%.

Several classifier evaluation performance indicators were used, such as accuracy (Ac), recall (Re), specificity (Sp), precision (Pr), sensitivity (Se), and F1 score. The classification performance of our proposed method is shown in Table 7.

Table 7. Classification performance of super vector machine classifier.

Trial	TP (%)	FP (%)	TN (%)	FN (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Recall (%)	Precision (%)	F1 Score (%)
Normal	100	0	82.6	17.4	91.30	85.17	100.00	85.17	100.00	91.99
Diabetes	82.6	17.4	100	0	91.30	100.00	85.17	100.00	82.60	90.47

3.4. Final Results

After classifying all the input data, the final results of the comparison of the three inputs were obtained, as shown in the Table 8. We conducted a comparison between the outcomes of our approach and those used in earlier research. A performance comparison with past investigations is shown in Table 9.

Table 8. Signal input comparison.

Data Input	Classification Model	Accuracy (%)	F1 Score (%)	Feature Points	Exercise Time (s)
Original PPG	Weighted KNN	86.96	88.46	2100	110
Instantaneous Frequency PPG	Super Vector Machine	89.15	90.21	63	1.67
Spectral Entropy PPG	Super Vector Machine	91.30	91.99	63	9.25

Table 9. Classification performance comparison.

Year	PPG Signal	Invasive Methods	Classifier	Features Extraction	Evaluation Metric	Characteristic
2009	Finger sensor, H. Karimipour et al. [45]	Not mentioned	Auto-Regressive Moving Average (ARMA)	24,750	Sensitivity = 100%	Classification
2017	Pulse Oximeter, E. M. Moreno et al. [46]	HbA1c Test	Random Forest	9	ROC = 0.7	Classification
			Gradient Boosting	9	ROC = 0.7	Classification
			Linear Discriminant Analysis	9	ROC = 0.6	Classification
2019	Smartphone Camera, Y. Zhang et al. [47]	Glucose meter	Subspace KNN	67	Accuracy = 86.2%	Classification
			RUS Boosted Trees	67	Accuracy = 85.0%	Classification
			Bagged Trees	67	Accuracy = 86.0%	Classification
			Decision Trees	67	Accuracy = 80.1%	Classification
2020	Smartphone, G. Zhang et al. [21]	Glucose meter	Gaussian Super Vector Machine (GSVM)	28	Accuracy = 81.5%	Classification
			Bagged Trees	28	Accuracy = 74.0%	Classification
			K-Nearest Neighbor	28	Accuracy = 71.0%	Classification
2022	Finger sensor	Glucose meter	Weighted KNN	2100	Accuracy = 86.96%	Classification
			Super Vector Machine(Proposed method in this study)	63	Accuracy = 91.3%	Classification

4. Discussion

The signal is not a single item, but a set of sinusoidal components. The notion of frequency loses all significance when applied to a non-stationary signal, making it essential to use the idea of time-varying process parameters to accurately estimate the sinusoidal transmission frequency.

Similar to many other biological signals, PPG signals are non-stationary. The amplitude, frequency, and oscillation pattern of the PPG impulse change from beat to beat. Its bandwidth and frequency may fluctuate over time. Non-stationary approaches, such as frequency analysis, are necessary for non-stationary signals [43].

A useful method for comprehending a signal's non-stationarity is time–frequency (TF) analysis. In this stage, a spectrogram is utilized to visually evaluate the form of the PPG signal based on its BGL classification and to identify the amount of noise it contains.

Each PPG signal has essential data that are relevant to the degree of precision of the outputs of the machine learning classification procedure in the form of feature points. However, if there are too many feature points, training will take a long time. The original ppg signal has 2100 points and takes roughly 110 s to train, while IF and SE only have 63 points and take less than 10 s to train. To calculate the spectrogram, IF and SE run windows 63 times. The number of feature points in each data point was decreased from 2100 lengthy signal samples to only 63. The training procedure could take just a few seconds with a modest training set.

The final findings are shown in Table 8, and they demonstrate that the input signal that has been processed by TF delivers a higher degree of accuracy. This result further shows the need of signal pre-processing for the PPG signal. In earlier investigations, the steps used during signal pre-processing were mainly focused on removing noise that was

built into the PPG. This work establishes that classification accuracy may be improved without noise elimination by converting time-based PPG through the short-time Fourier transform (STFT).

The average of the instantaneous frequency and the spectral entropy differs by almost an order of magnitude. IF and SE signal types are identical in diabetic individuals, as can be seen from Figure 19, but the amplitude values should be noted, since IF has a greater amplitude value. Due to the average instantaneous frequency being too high to investigate machine learning successfully, SE is thus more accurately based on the classification findings. SE is seen based on the significant change in the spectrum of successive frames. Big inputs may create network convergence when the network fits data with a large average and a wide range of values, which can impact the classification accuracy level.

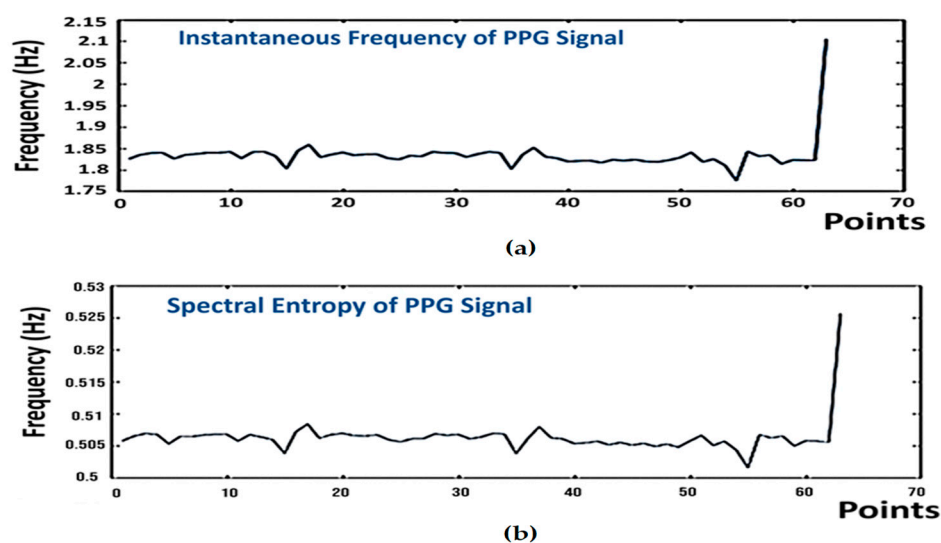


Figure 19. IF and SE signals in diabetic individuals: (a) IF PPG signal; (b) SE PPG signal.

The process of selecting the machine learning algorithm that best matches the non-linear nature of the PPG signal follows signal pre-processing. The accuracy of the outputs, overall cost of misclassification, and training time were factors in the algorithm selection. The top three were discovered after a series of comparisons between several machine learning classifiers using the same dataset. The super vector machine (SVM), with 91.3% accuracy and a training duration of 9 s, was the best classifier. Based on these findings, we decided to use non-invasive PPG and the SVM to categorize BGL.

We conducted a comparison between the outcomes of our approach and those of earlier research. The comparison study's findings demonstrate that the model we suggest produces a superior accuracy score and overall outcome. In general, although using the same classifier, outcomes from quantitative research in earlier studies might vary, as seen in Table 9. We concentrated on comparing the findings to those of earlier scholars. The Gaussian SVM approach was proposed by G. Zhang et al. [21] with a final accuracy rating of 81.5%. The SVM algorithm is taught to split the input data into two groups that are separated by as much space as feasible. The SVM classifier is sometimes referred to as a the broad margin classifier because of the wide space between these category labels. The SVM model, however, is a linear classifier and is unable to map nonlinear functions without the use of kernels. Because it employs the same technique as our suggested approach, the results of G. Zhang et al. offer a unique comparison. However, these findings are less accurate than those from our suggested model. Due to variations in the PPG raw signal's quality and the amount of sample points utilized, the accuracy numbers change.

5. Conclusions

In comparison to previous BGL estimate techniques, this categorization technique provides benefits. With our suggested procedure, users may quickly determine the state of their BGL to guarantee early diabetes identification. This technique may reduce the chance of mortality while accelerating the healing process. A more ideal accuracy value of machine learning classification results will be attained by retaining PPG signal quality from the start. Using our suggested technique, it is not necessary to extract the morphological elements of the PPG signal, making it applicable in a wide range of circumstances. Compared to decision trees, subspace k-nearest neighbor, and k-nearest neighbor, our suggested approach with SVM obtains a greater level of accuracy. With an overall F1 score of 91.99% across the classification levels, the SVM classifier has the greatest overall score. Further enhancing the effectiveness of BGL categorization based on PPG signals may be conducted by using larger sample sizes. We will employ an embedded system in a future study to combine the machine learning algorithms with electrical circuitry to produce portable monitoring devices.

Author Contributions: E.S. designed the methodology and software, data collection; K.R. Advised and supervised the project and corresponding; P.D.P. Advised and supervised the experiments and analysis; N.H.A. Supervised data processing, editing the paper; and all authors contributed to writing the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by PDD grant of Ristekdikti, grant number: NKB-1010/UN2.RST/HKP.05.00/2022.

Data Availability Statement: Data available in a publicly accessible repository. The data presented in this study are openly available in [Figshare] at [<https://doi.org/10.6084/m9.figshare.5459299>], [41].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. IDF DIABETES ATLAS, Seventh Edition. 2015. Available online: <https://www.idf.org/e-library/welcome.html> (accessed on 1 February 2023).
2. Ahuja, A.; Gupta, R. Diabetes Silent Killer: Medical focus on Food Replacement and Dietary Plans. *Adv. Bioresarch* **2020**, *11*, 128–135. [[CrossRef](#)]
3. Carracher, A.M.; Marathe, P.H.; Close, K.L. International Diabetes Federation 2017. *J. Diabetes* **2018**, *10*, 353–356. [[CrossRef](#)] [[PubMed](#)]
4. Banoo, H.; Nusrat, N.; Nasir, N. Type2 diabetes mellitus: A review of current trends. *RAMA Univ. J. Med. Sci.* **2015**, *1*, 50–57.
5. American Diabetes Association 2. Classification and Diagnosis of Diabetes. *Diabetes Care* **2015**, *38*, S8–S16. [[CrossRef](#)]
6. WHO. *Global Report on Diabetes*; WHO: Geneva, Switzerland, 2016.
7. Nawaz, A.; Øhlckers, P.; Sælid, S.; Jacobsen, M.; Akram, M.N. Review: Non-Invasive Continuous Blood Glucose Measurement Techniques. *J. Bioinform. Diabetes* **2016**, *1*, 1–27. [[CrossRef](#)]
8. Nicholas, D.; Logan, K.A.; Sheng, Y.; Gao, J.; Farrell, S.; Dixon, D.; Callan, B.; McHale, A.P.; Callan, J.F. Rapid paper based colorimetric detection of glucose using a hollow microneedle device. *Int. J. Pharm.* **2018**, *547*, 244–249. [[CrossRef](#)] [[PubMed](#)]
9. Galaviz, K.I.; Narayan, K.M.V.; Lobelo, F.; Weber, M.B. Lifestyle and the Prevention of Type 2 Diabetes: A Status Report. *Am. J. Lifestyle Med.* **2015**, *12*, 4–20. [[CrossRef](#)]
10. Alsunaidi, B.; Althobaiti, M.; Tamal, M.; Albaker, W.; Al-Naib, I. A Review of Non-Invasive Optical Systems for Continuous Blood Glucose Monitoring. *Sensors* **2021**, *21*, 6820. [[CrossRef](#)]
11. Villena Gonzales, W.; Mobashsher, A.T.; Abbosh, A. The Progress of Glucose Monitoring—A Review of Invasive to Minimally and Non-Invasive Techniques, Devices and Sensors. *Sensors* **2019**, *19*, 800. [[CrossRef](#)]
12. Tang, L.; Chang, S.J.; Chen, C.-J.; Liu, J.-T. Non-Invasive Blood Glucose Monitoring Technology: A Review. *Sensors* **2020**, *20*, 6925. [[CrossRef](#)]
13. Cengiz, E.; Tamborlane, W.V. A Tale of Two Compartments: Interstitial Versus Blood Glucose Monitoring. *Diabetes Technol. Ther.* **2009**, *11*, S11. [[CrossRef](#)] [[PubMed](#)]
14. Monte-Moreno, E. Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. *Artif. Intell. Med.* **2011**, *53*, 127–138. [[CrossRef](#)] [[PubMed](#)]
15. Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **2007**, *28*, R1–R39. [[CrossRef](#)]
16. Tjahjadi, Ramli, K. Noninvasive Blood Pressure Classification Based on Photoplethysmography Using K-Nearest Neighbors Algorithm: A Feasibility Study. *Information* **2020**, *11*, 93. [[CrossRef](#)]

17. Pilt, K.; Meigas, K.; Ferenets, R.; Temitski, K.; Viigimaa, M. Photoplethysmographic signal waveform index for detection of increased arterial stiffness. *Physiol. Meas.* **2014**, *35*, 2027–2036. [CrossRef] [PubMed]
18. Lei, R.; Ling, B.W.-K.; Feng, P.; Chen, J. Estimation of Heart Rate and Respiratory Rate from PPG Signal Using Complementary Ensemble Empirical Mode Decomposition with both Independent Component Analysis and Non-Negative Matrix Factorization. *Sensors* **2020**, *20*, 3238. [CrossRef] [PubMed]
19. Sen Gupta, S.; Kwon, T.-H.; Hossain, S.; Kim, K.-D. Towards non-invasive blood glucose measurement using machine learning: An all-purpose PPG system design. *Biomed. Signal Process. Control* **2021**, *68*, 102706. [CrossRef]
20. Castaneda, D.; Esparza, A.; Ghamari, M.; Soltanpur, C.; Nazeran, H. A review on wearable photoplethysmography sensors and their potential future applications in health care. *Int. J. Biosens. Bioelectron.* **2018**, *4*, 195–202. [CrossRef]
21. Zhang, G.; Mei, Z.; Zhang, Y.; Ma, X.; Lo, B.; Chen, D.; Zhang, Y. A Noninvasive Blood Glucose Monitoring System Based on Smartphone PPG Signal Processing and Machine Learning. *IEEE Trans. Ind. Inform.* **2020**, *16*, 7209–7218. [CrossRef]
22. Tyapochkin, K.; Smorodnikova, E.; Pravdin, P. Smartphone PPG: Signal processing, quality assessment, and impact on HRV parameters. 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. [CrossRef]
23. Caizzone, A.; Boukhayma, A.; Enz, C. AC/DC Ratio Enhancement in Photoplethysmography Using a Pinned Photodiode. *IEEE Electron Device Lett.* **2019**, *40*, 1828–1831. [CrossRef]
24. Caizzone, A.; Boukhayma, A.; Enz, C. Comprehensive noise analysis in PPG read-out chains. In Proceedings of the 2017 International Conference on Noise and Fluctuations (ICNF), Vilnius, Lithuania, 20–23 June 2017.
25. Nara, S.; Kaur, M.; Verma, K.L. Novel Notch Detection Algorithm for Detection of Dicrotic Notch in PPG Signals. *Int. J. Comput. Appl.* **2014**, *86*, 36–39. [CrossRef]
26. Serra, C.; Sestu, A.; Murru, V.; Greco, G.; Vacca, M.; Scuteri, A. Diabetes Affects the Relationship between Heart Rate Variability and Arterial Stiffness in a Gender-Specific Manner. *J. Clin. Med.* **2022**, *11*, 4937. [CrossRef] [PubMed]
27. ChuDuc, H.; NguyenPhan, K.; NguyenViet, D. A Review of Heart Rate Variability and its Applications. *APCBEE Procedia* **2013**, *7*, 80–85. [CrossRef]
28. Benichou, T.; Pereira, B.; Mermillod, M.; Tauveron, I.; Pfabigan, D.; Maqdasy, S.; Dutheil, F. Heart rate variability in type 2 diabetes mellitus: A systematic review and meta-analysis. *PLoS ONE* **2018**, *13*, e0195166. [CrossRef] [PubMed]
29. Lu, G.; Yang, F.; Taylor, J.; Stein, J.F. A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. *J. Med. Eng. Technol.* **2009**, *33*, 634–641. Available online: <https://www.tandfonline.com/doi/full/10.3109/03091900903150998?scroll=top&needAccess=true> (accessed on 1 February 2023). [CrossRef]
30. Selvaraj, N.; Jaryal, A.; Santhosh, J.; Deepak, K.K.; Anand, S. Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *J. Med Eng. Technol.* **2008**, *32*, 479–484. [CrossRef]
31. Rothberg, L.; Lees, T.; Clifton-Bligh, R.; Lal, S. Association Between Heart Rate Variability Measures and Blood Glucose Levels: Implications for Noninvasive Glucose Monitoring for Diabetes. *Diabetes Technol. Ther.* **2016**, *18*, 366–376. [CrossRef]
32. Alian, A.A.; Shelley, K.H. Photoplethysmography. *Best Pr. Res. Clin. Anaesthesiol.* **2014**, *28*, 395–406. [CrossRef]
33. Vahlsing, T.; Delbeck, S.; Leonhardt, S.; Heise, H.M. Noninvasive Monitoring of Blood Glucose Using Color-Coded Photoplethysmographic Images of the Illuminated Fingertip Within the Visible and Near-Infrared Range: Opportunities and Questions. *J. Diabetes Sci. Technol.* **2018**, *12*, 1169–1177. [CrossRef]
34. Chellappan, K.; Sahathevan, R. Accelerated Aging and Noninvasive Cardiovascular Risk Monitoring. *Cerebrovasc. Dis.* **2015**, *40*, 25–27.
35. Spigulis, J.; Kukulis, I.; Fridenberga, E.; Venckus, G. Potential of advanced photoplethysmography sensing for non-invasive vascular diagnostics and early screening. In *The International Society for Optical Engineering*; Society of Photo-Optical Instrumentation Engineers (SPIE) Press: San Jose, CA, USA, 2002; Volume 4862, pp. 38–43. [CrossRef]
36. Hettiarachchi, C.; Chitraranjan, C. A Machine Learning Approach to Predict Diabetes Using Short Recorded Photoplethysmography and Physiological Characteristics. In Proceedings of the 17th Conference on Artificial Intelligence in Medicine, Poznan, Poland, 26–29 June 2019; Science, S.C., Ed.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 322–327. [CrossRef]
37. Qawqzeh, Y.K.; Bajahzar, A.S.; Jemmali, M.; Otoom, M.M.; Thaljaoui, A. Classification of Diabetes Using Photoplethysmogram (PPG) Waveform Analysis: Logistic Regression Modeling. *BioMed Res. Int.* **2020**, 3764653. [CrossRef] [PubMed]
38. Hossain, S.; Debnath, B.; Biswas, S.; Al-Hossain, M.J.; Anika, A.; Navid, S.K.Z. Estimation of Blood Glucose from PPG Signal Using Convolutional Neural Network. In Proceedings of the 2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health, Qingdao, China, 28–30 November 2019. [CrossRef]
39. Deng, H.; Zhang, L.; Mo, Y.X.S. Research on Estimation of Blood Glucose Based on PPG and Deep Neural Networks. In Proceedings of the 8th Annual International Conference on Geo-Spatial Knowledge and Intelligence, Xi'an, China, 18–19 December 2020. [CrossRef]
40. Liang, Y.; Chen, Z.; Liu, G.; Elgendi, M. A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China. *Sci. Data* **2018**, *5*, 180020. [CrossRef] [PubMed]
41. Liang, Y.; Liu, G.; Chen, Z.; Elgendi, M. PPG-BP Database. Figshare. Dataset. Available online: https://figshare.com/articles/dataset/PPG-BP_Database_zip/5459299/5 (accessed on 5 May 2022).
42. Elgendi, M. Optimal Signal Quality Index for Photoplethysmogram Signals. *Bioengineering* **2016**, *3*, 21. [CrossRef]

43. Stanković, L.; Djurović, I.; Stanković, S.; Simeunović, M.; Djukanović, S.; Daković, M. Instantaneous frequency in time–frequency analysis: Enhanced concepts and performance of estimation algorithms. *Digit. Signal Process.* **2014**, *35*, 1–13. [\[CrossRef\]](#)
44. Romulus, R. A Comparison Between Instantaneous frequency Estimation Methods of Frequency Modulated Signals Covered with Gaussian Noise. In Proceedings of the 2012 10th International Symposium on Electronics and Telecommunications, Timisoara, Romania, 15–16 November 2012. [\[CrossRef\]](#)
45. Karimipour, H.; Shandiz, H.T.; Zahedi, E. Diabetic diagnose test based on PPG signal and identification system. *J. Biomed. Sci. Eng.* **2009**, *2*, 465–469. [\[CrossRef\]](#)
46. Moreno, E.M.; Luján, M.J.A.; Rusiñol, M.T.; Fernández, P.J.; Manrique, P.N.; Triviño, C.A.; Miquel, M.P.; Rodríguez, M.A.; Burguillos, M.J.G. Type 2 Diabetes Screening Test by Means of a Pulse Oximeter. *IEEE Trans. Biomed. Eng.* **2016**, *64*, 341–351. [\[CrossRef\]](#)
47. Zhang, Y.; Zhang, Y.; Siddiqui, S.; Kos, A. Non-invasive blood-glucose estimation using smartphone PPG signals and subspace kNN classifier. *Elektrotehniski Vestn.* **2019**, *86*, 68–74.

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.