Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 6 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i06.53815

PAPER

A Stratified Modeling-Machine Learning Approach to Improve the Accuracy of Non-Invasive Blood Glucose Estimation Using Photoplethysmography Signals

Fitrilina[1,2], Muhammad Ilhamdi Rusydi[1](✉), Rahmadi Kurnia[1], Noverika Windasari[3], Raffi Zahrandhika Putra[2]

[1]Department of Electrical Engineering, Faculty of Engineering, Universitas Andalas, Padang City, Indonesia

[2]Department of Electrical Engineering, Faculty of Engineering, Universitas Bengkulu, Bengkulu City, Indonesia

[3]Department of Forensic and Legal Medicine, Faculty of Medicine, Universitas Andalas, Padang City, Indonesia

rusydi@eng.unand.ac.id

ABSTRACT

Diabetes is a silent killer that can only be controlled with continuous monitoring of blood glucose levels. The method commonly used is invasive and has various weaknesses, but it is more accurate than non-invasive methods. This research aims to develop a method to increase the accuracy of non-invasive estimation of blood glucose levels using photoplethysmography (PPG) signals. The proposed method is to carry out stratified modeling-machine learning. The tested classifiers were support vector machines (SVM), KNN, Naïve Bayes, decision tree, and neural network. The prediction model used simple linear, logarithmic, second-order polynomial, exponential, and power regression. Applying stratified modeling using linear regression in the non-diabetes stratum and logarithmic regression in the diabetes stratum obtained a mean absolute relative difference (MARD) value of 4.5%, root mean square error (RMSE) of 18.9 mg/dl, Pearson correlation 0.985 and Clarke error grid analysis (CEGA) 96% in region A and 4% in region B. The implementation of stratification reveals a marked improvement in efficacy, manifested as a reduction in the MARD by 77.83%, a decrease in the RMSE by 51.91%, an enhancement in the Pearson correlation by 0.065, and a CEGA by 100% in regions A and B, thereby being clinically acceptable. Implementing a stratified modeling-machine learning approach can improve the accuracy of non-invasive blood glucose level estimates.

KEYWORDS

blood glucose non-invasively, machine learning, photoplethysmography (PPG), classification, regression

1 INTRODUCTION

The International Diabetes Federation (IDF) identifies diabetes as one of the top ten causes of adult mortality. In 2017, approximately 424.9 million adults (aged 20–79 years) were diagnosed with diabetes, resulting in 4 million deaths and global

Fitrilina, Rusydi, M.I., Kurnia, R., Windasari, N., Putra, R.Z. (2025). A Stratified Modeling-Machine Learning Approach to Improve the Accuracy of Non-Invasive Blood Glucose Estimation Using Photoplethysmography Signals. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(6), pp. 76–96. https://doi.org/10.3991/ijoe.v21i06.53815

Article submitted 2024-12-12. Revision uploaded 2025-01-30. Final acceptance 2025-01-30.

healthcare expenditures of USD 727 billion [1]. By 2019, these numbers increased to 463 million cases, 4.2 million deaths, and USD 760.3 billion in expenditures. The global prevalence rose from 8.8% in 2017 to 9.3% in 2019, with projections indicating a continued upward trend. In 2019, Indonesia ranked seventh globally and was the only Southeast Asian country among the top ten nations with the highest number of diabetes cases in the 20–79 age group [1], [2], and [3].

Diabetes is a chronic disease [4], a metabolic disorder [5], [6], characterized by fluctuations in blood glucose levels [6], and is a lifelong disease [7]. Diabetes is characterized by impaired insulin production and utilization, leading to elevated blood glucose levels and affecting digestion [8]. Uncontrolled hyperglycemia, with glucose levels above the euglycemic range [9], can cause severe damage to the heart, blood vessels, nerves, eyes, and kidneys [5], [6]. While hypoglycemia, in which glucose levels fall below the euglycemic range [9], can cause seizures, coma, arrhythmias, and cardiac failure [6]. Effective diabetes management relies on continuous blood glucose monitoring. Current methods are invasive or minimally invasive, requiring punctured blood samples, causing pain, discomfort, and high cost [10], [11]. This method uses a disposable test strip and places a drop of blood sample for analysis. Blood glucose monitoring for patients with glycemic disorders can consume up to 25% of their healthcare costs, primarily due to the high expense of glucose test strips, which dominate 85% of the global biosensor market [12]. Therefore, non-invasive methods are crucial to addressing the limitations of invasive methods [13].

There are many studies, but no reliable and accurate method has been established [10], [11], [14], [15], and [16] which requires further improvement [10]. Non-invasive blood glucose screening is categorized into optical and non-optical methods. The optical properties of glucose are specific and show a superior correlation. Thus, optical methods provide better results than non-optical methods [17]. Glucose concentration affects the scattering coefficient, thus allowing estimation through light intensity changes transmitted or reflected from glucose-containing tissues [6], [13]. A typical optical method used for measuring blood glucose concentration is photoplethysmography (PPG) [3], [17], and [18]. PPG is a noninvasive methodology that measures blood volume by examining the intensity of light transmitted or reflected from human biological tissues [19]. Previous research has shown that this technique can evaluate important physiological metrics such as blood pressure, heart rate, and respiration while identifying diabetes [19].

Research into methods for non-invasive estimation of glucose levels has been extensively advanced. The most rudimentary approach entails mathematical modeling, as evidenced in studies [4] and [10]. Research [20], [21] employs statistical methodologies. The most widely utilized form of simple regression is linear regression [4], [6], [13], [21], [22], [23], and [24], alongside polynomial regression [7], [18], and [22]. Conventional methods are widely utilized due to their inherent simplicity; nevertheless, the accuracy and robustness of these methods require further refinement. The machine learning techniques that are commonly employed include partial least squares (PLS) [20], [25]; random forest (RF) [26], [27], and [28]; ensembled boosted trees-SVR [29]; the ensemble regression trees model [30]; and the ensemble bagging tree [3], in addition to ANN-based frameworks [12], [31]; support vector machines (SVM) [32]; FGSVR [33]; exponential Gaussian process machine learning regression [34]; Kernel-based regression methods [35]; and multiple linear regression analysis [36]. Applying non-linear machine learning methods yields superior estimation outcomes compared to linear machine learning approaches [5]. This research focused on developing prediction models for blood glucose estimation using data from diabetic and non-diabetic subjects. The training process combines these datasets into a single model, introducing significant heterogeneity due to variations in PPG signal patterns between diabetic and non-diabetic groups [3] and individual differences within each group [3], [37]. Therefore, combining diabetic data and non-diabetic data in one group of training data sets will increase the heterogeneity of the data set. Heterogeneity refers to the variability present within data sets, requiring thorough examination to mitigate potential erroneous outcomes and conclusions [37], [38], and [39].

Data heterogeneity is a challenge in prediction or estimation models because data is very diverse, it is difficult to capture consistent patterns, and it is more complex [37], [40] because more factors must be considered, thus increasing estimation bias and reducing machine learning accuracy [37]. This study addresses these issues by proposing a noninvasive blood glucose estimation approach using homogeneous data groups. By aligning the model with the specific characteristics of the data, this approach aims to reduce variability, enhance model precision, and simplify the decision-making process. This study's contribution lies in the implementation of stratified modeling to estimate blood glucose levels noninvasively through the utilization of PPG signals in conjunction with machine learning methods. This study aims to enhance the accuracy of the noninvasive measurement of blood glucose concentrations by implementing stratified modeling techniques. This study will evaluate system accuracy with and without stratified modeling methods.

2 MATERIALS AND METHODS

This study employs stratified modeling with machine learning to estimate blood glucose levels non-invasively, as shown in Figure 1. Stratified modeling involves developing separate models for distinct categorical features or parameter values [41]. Segmenting heterogeneous datasets into homogeneous clusters simplifies data distribution and facilitates analysis using conventional methods [37]. Models built within these clusters better align with the data's characteristics, reducing variability and enhancing predictive accuracy. The effectiveness of the stratified approach relies heavily on selecting stratification features, effectively balancing the strengths of complex and simple models [41].

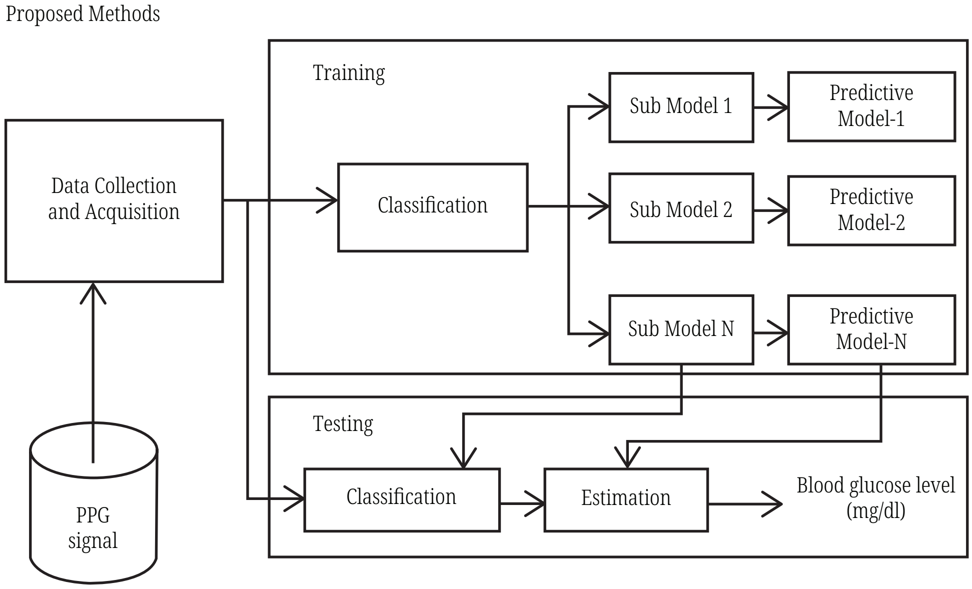


Fig. 1. Stratified modelling process to estimate blood glucose concentration

This study utilized diabetes and non-diabetes as stratification parameters, employing a machine learning classifier to categorize the data. Separate predictive models were developed for each stratum. During training, the classifier was trained to distinguish between diabetic and non-diabetic data while simultaneously building predictive models specific to each category. In testing, PPG signals were classified as diabetic or non-diabetic, and blood glucose concentrations were estimated using the corresponding predictive model. The research workflow is summarized in Figure 2.

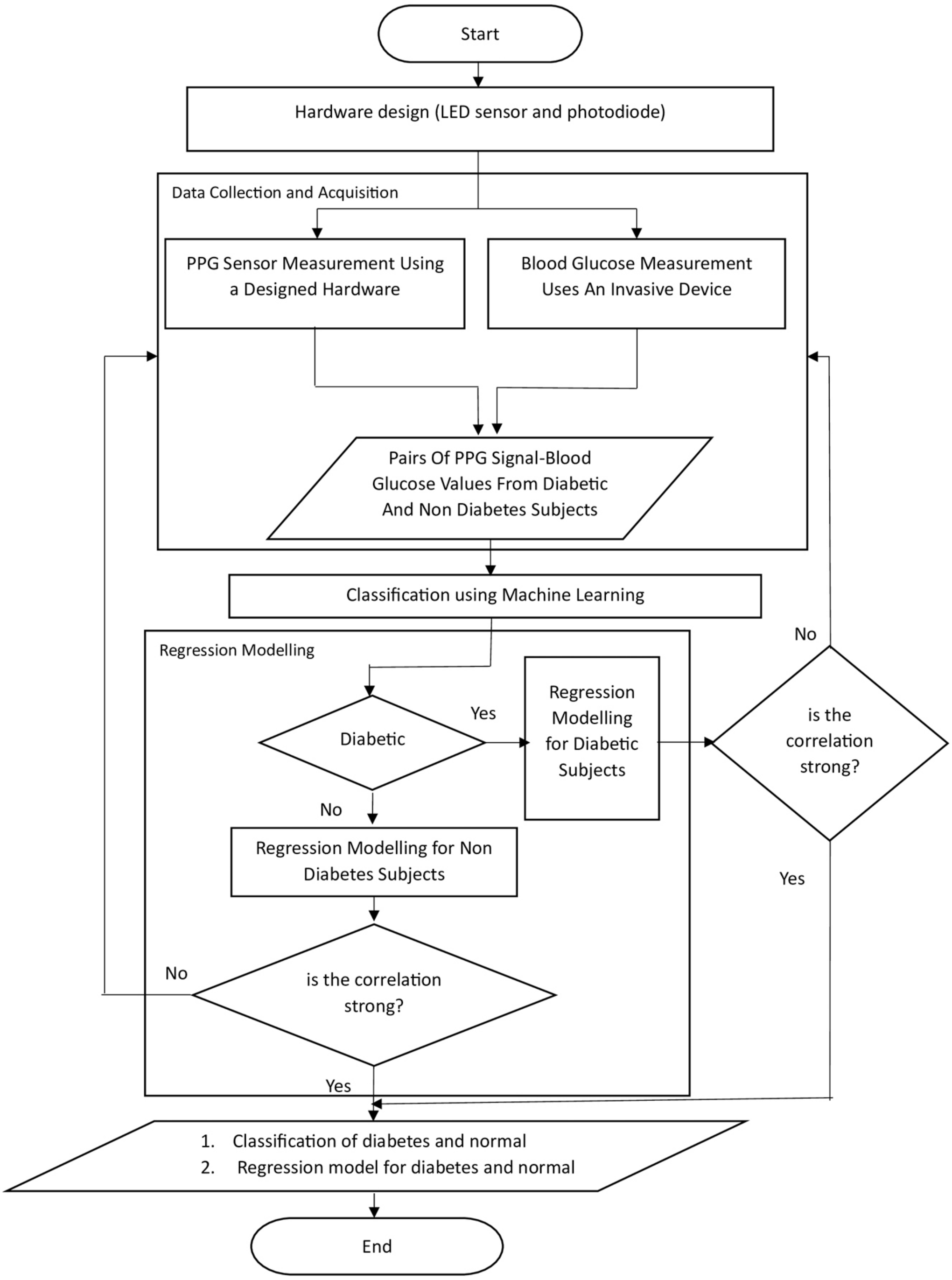


Fig. 2. Flowchart of research stages with a stratified modelling approach

2.1 Hardware design

Two sensor positioning modes can be employed: the transmission mode (see Figure 3a) and the reflectance mode (see Figure 3b).

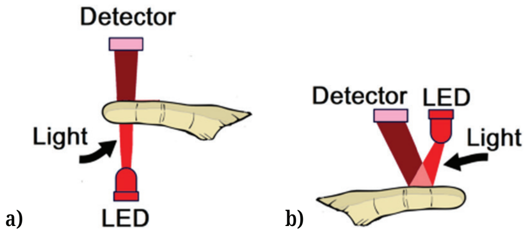


Fig. 3. Sensor mode: a) Transmission, b) Reflectance [18]

This study utilized the transmission mode. Near-infrared light-emitting diodes (NIR-LEDs) operating at 940 nm emit light that is absorbed and converted into voltage signals by the OPT101. The wavelength of 940 nm is selected due to its high sensitivity to glucose molecules in the bloodstream [4], [6], and [16]. The PPG signals were acquired using the OPT101 sensor. The OPT101 was chosen for its high sensitivity and ability to minimize interference from environmental light or scattering effects [42]. Its integrated photodiode detects light intensity, which is converted into a current output. The trans-impedance amplifier (TIA) within the OPT101 amplifies this current and converts it into a proportional voltage output, directly correlating with blood glucose concentration [43]. No additional preprocessing techniques, such as filtering or noise reduction, were applied, relying instead on the sensor's ability to minimize interference from environmental light and signal artifacts. The raw signal was directly used for further analysis and modelling. The analog voltage is processed by the ESP32 to be converted to an ADC value.

2.2 Data collection and acquisitions

Photoplethysmography signals were recorded over 20 seconds at a sampling frequency of 100 Hz. The analog voltage was converted into ADC values ranging from 0 to 1023 (10-bit resolution). The data set consists of sensor voltage outputs (ADC value), subject ages, and blood glucose concentrations (mg/dL) obtained from an invasive glucometer (Accu-Chek). PPG signals were labeled as non-diabetes or diabetic based on the 2019 National Institute for Clinical Excellence (NICE) guidelines, as detailed in Table 1.

Table 1. Blood glucose level of nice 2019 [3]

Plasma Glucose Test	Non-Diabetes	Prediabetes	Diabetes
Random	<200 mg/dl or <11,1 mmol/L	N/A	>200 mg/dl or >11,1 mmol/L
Fasting	<100 mg/dl or <5,5 mmol/L	100–125 mg/dl or 5,5–6,9 mmol/L	>126 mg/dl or 7 mmol/L

Random plasma glucose assessments may be performed at any time without the necessity for preparatory measures [3]. Conversely, the fasting plasma glucose evaluation is conducted following a minimum fasting period of eight hours [3]. This research employed a random testing methodology. The study analyzed 200 training datasets comprising 100 non-diabetes and 100 diabetic samples and 50 testing data sets comprising 25 non-diabetes and 25 diabetic samples. This study received ethical approval from the Ethics Committee of the Faculty of Medicine, Universitas Andalas, with approval number 519/UN.16.2/KEP-FK/2024.

2.3 Classification

Classification was performed to categorize PPG signals into non-diabetes or diabetic strata. The classified ADC voltage values were then converted to blood glucose levels (mg/dL) using the corresponding regression model. Recently, machine learning methods are widely used in the medical field for the detection, diagnosis, and classification of diseases [44]. All models tested are shown in Figure 4.

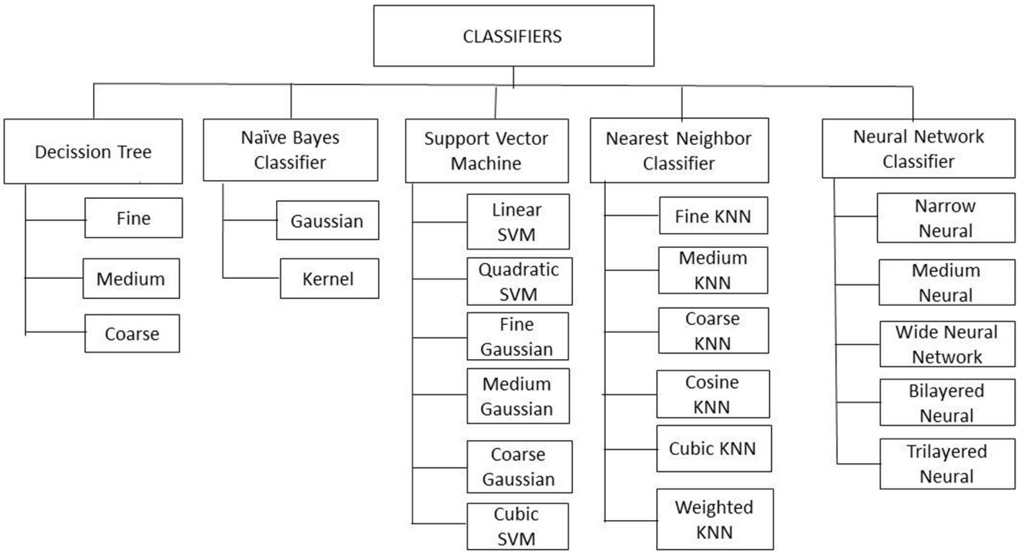


Fig. 4. Classifier tested

The validation process in this study employs both resubstitution and five-fold cross-validation techniques to evaluate model performance. Resubstitution validation provides an initial assessment of how well the model fits the training data, while five-fold cross-validation ensures robustness by partitioning the dataset into five subsets, using one subset for testing and the remaining for training iteratively. This approach minimizes overfitting and provides a more reliable estimate of the model's generalization capability.

2.4 Estimation

This study employs simple regression methods: linear, second-order polynomial, exponential, logarithmic, and power models. In this approach, the ADC value is the independent variable (*X*), and the invasive blood glucose measurement from the ACCU-CHECK glucometer is the dependent variable (*Y*). The efficacy of these estimation models was evaluated based on the following criteria:

1. Mean absolute relative difference (MARD): Mean absolute relative difference is a prevalent criterion utilized for assessing glucose monitoring methods [28], [31]. It quantifies the relative difference between a non-invasive device's predicted values and an invasive device's reference values [28], [31]. *MARD* is calculated using equation (1) [33].

$$MARD = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(X_i - Y_i)}{Y_i}\right| \times 100\ \% \tag{1}$$

2. Root mean square error (RMSE): Root mean square error measures the deviation between predicted and actual values, reflecting the standard deviation of residuals [28], [31]. It is calculated using equation (2) [28].

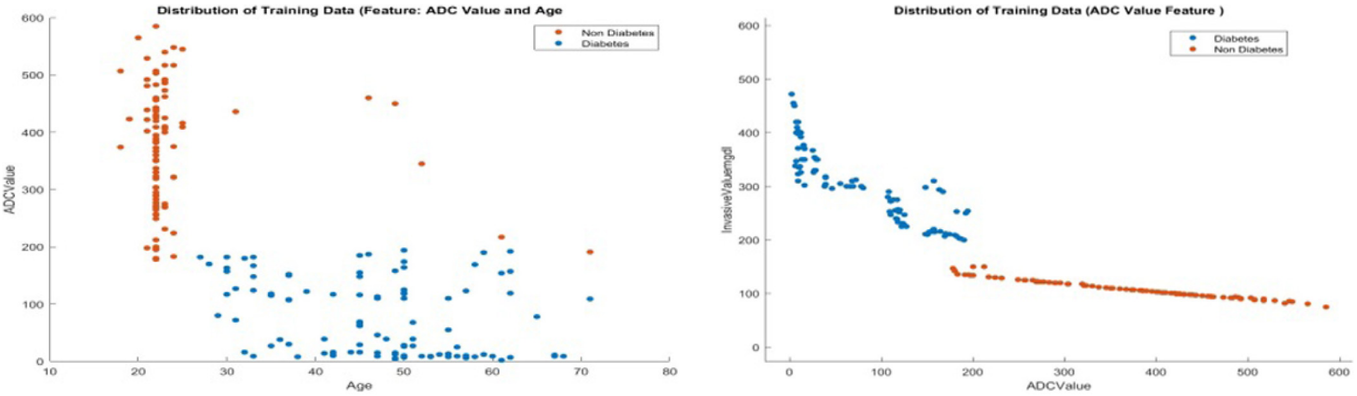
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{n}} \tag{2}$$

   x denotes the predicted value non-invasively, y represents the reference value invasively, and n indicates the number of samples.
3. Pearson correlation (R): Correlation coefficient (R) describes the strength of the relationship [45]. This study uses Pearson correlation to analyze the relationship between the estimated blood glucose values from the regression model and the reference values measured by an invasive glucometer.
4. Clark error grid analysis (CEG): Clarke error grid analysis (CEGA) is a standard method for assessing the clinical accuracy of self-monitoring of blood glucose (SMBG) devices [46]. Region A is clinically accurate decisions (deviation of ±20% from the actual values) [47]; Region B is clinically acceptable and uncritical decisions; Region C is overcorrections that could lead to an adverse outcome; Region D is dangerous failure to detect and treat; and Region E is erroneous treatment [48].

3 RESULTS AND DISCUSSION

3.1 Data acquisition and analysis

The dataset in this study, while demonstrating the methodology's efficacy with participants from Bengkulu City, Indonesia (93 male and 107 female, aged 18–71 years), does not include key demographic variables such as ethnicity, body mass, and pre-existing conditions. Although this limits generalizability to diverse populations, the primary objective was to evaluate the effectiveness of a stratified modelling approach for improving non-invasive glucose estimation, which was successfully achieved within this context. Future research should build on these findings by incorporating more diverse datasets with broader demographic and physiological factors to enhance the model's robustness and ensure clinical utility across varied populations. Figure 5 shows the data set distribution used in the training phase. The ADC value data and the age of the subjects can be seen in Figure 5a. The distribution of the output voltage value of the OPT101 sensor (in ADC units) corresponding to the blood glucose level measured by the Accu-Chek device is illustrated in Figure 5b.



a) Two features

b) One feature

Fig. 5. Distribution of ADC value in training data

Figure 5b shows an inverse correlation between sensor voltage and blood glucose concentration, where an increase in glucose levels leads to a decrease in sensor voltage. This finding aligns with previous studies [49], suggesting that higher glucose concentrations raise the attenuation coefficient, causing greater light absorption and less scattering in the blood [50]. Consequently, less light is captured by the sensor, reducing the PPG signal amplitude and sensor voltage [13] [51]. Figure 5b shows distinct distribution patterns of ADC values and blood glucose levels between non-diabetes and diabetic individuals. The non-diabetes category exhibits a more linear distribution compared to the diabetic category. As a result, using a single predictive model, such as a linear regression equation, for all subjects is likely to introduce significant estimation bias, negatively affecting model accuracy [37]. To reduce this bias and improve accuracy, we propose developing separate predictive models for each category, non-diabetes and diabetic, through stratified modeling. This approach classifies subjects based on their PPG signals using various machine learning methods, followed by applying regression techniques for each group.

3.2 Classification results

The proposed method begins with a classification phase that distinguishes PPG signals into diabetic and non-diabetic categories. This step is crucial, as grouping the PPG signals enables more accurate subsequent estimations using an appropriate predictive model, reducing bias and improving precision. The testing dataset includes 50 samples (25 diabetic and 25 non-diabetic), with two classification schemes: one using both ADC value and age and the other using only the ADC value. This design evaluates whether age affects the classification or if the ADC value is the primary factor. Classification accuracy results are summarized in Table 2.

Table 2. Percentage accuracy of the classification process

Method	Using ADC Voltage and Age Features				Using ADC Voltage Features and 5-Fold Cross-Validation	
	Resubstitution Validation		5-Fold Cross Validation			
	Training	Testing	Training	Testing	Training	Testing
SVM:						
Linear	98.5	100	97.5	100	95	98
Quadratic	99.5	100	99	100	97.5	100
Cubic	99.5	100	98.5	100	97.5	100
Fine Gaussian	99.5	100	97.5	100	96.5	100
Medium Gaussian	99	100	98.5	100	95	100
Coarse Gaussian	97	100	97	100	95	100
KNN:						
Fine	100	100	97.5	100	99.5	100
Medium	98.5	100	97.5	100	97.5	100
Coarse	97	100	97	100	93.5	94
Cosine	98.5	100	97	100	93.5	94
Cubic	98.5	100	97.5	100	97.5	100
Weighted	100	100	98	100	99.5	100
Naïve Bayes:						
Gaussian	97.5	100	97	100	95	100
Kernel	99	100	99	100	95	98
Decision Trees:						
Fine	99.5	100	98.5	100	98	100
Medium	99.5	100	98.5	100	98	100
Coarse	99.5	100	98.5	100	98.5	100
Neural Network:						
Narrow	100	100	99.5	100	97	100
Medium	100	100	99.5	100	97	100
Wide	100	100	99.5	100	99.5	100
Bi-layered	100	100	99	100	99	100
Tri-layered	100	100	99.5	100	98	100

The classification results demonstrate that wide neural networks consistently outperformed other methods, achieving 100% accuracy in testing across all validation setups, highlighting its robustness. Models incorporating ADC voltage and age features showed significantly better performance than those using ADC voltage alone, underscoring the importance of combining signal-based and demographic data. Weighted KNN and quadratic SVM also delivered high accuracy, with performance ranging from 97.5% to 100%, while fine and medium decision trees proved reliable, achieving accuracy between 98% and 99.5%. Kernel Naïve Bayes showed strong results with 99% accuracy, outperforming its Gaussian variant. However, models using only ADC voltage showed limitations, particularly in KNN and SVM variants, reflecting the value of richer feature sets. These findings establish wide neural networks as the most robust approach, with feature selection playing a critical role in optimizing classification accuracy.

In this research, the classification process focuses on categorizing ADC values from PPG signals into diabetic or non-diabetic strata based on the 2019 NICE guidelines rather than diagnosing individuals as diabetic or non-diabetic. In contrast, previous studies on diabetes detection are summarized in Table 3.

Table 3. Related work

Authors	Method	Feature	Accuracy
Reddy et al. [52]	SVM	HRV	82%
Nirala et al. [53]	SVM	Signal and derivatives + eigen value	97.87%
Hettiarachchi et al. [54]	LDA	Signal feature + Physio data	83%
Qawqzeh et al. [55]	Logistic Regression	Signal feature + Physio data	92.30%
Prabha et al. [56]	SVM	MFCC + Physio data	92.28%
Chu et al. [57]	Logistic Regression	HRV + Physio data	90%
Susana et al. [3]	Ensemble Bagging Trees	2100 sample points	98%

Table 3 highlights previous studies that aimed to detect diabetes using various features of PPG signals, reporting high accuracies ranging from 82% to 98%. This study focuses on classifying the average voltage value of the PPG signal into diabetic and non-diabetic groups, a feature not explored in previous studies listed in Table 3. The average voltage of the PPG signal correlates with optical light intensity, which reflects glucose concentration in the blood. While the current approach demonstrates promising potential, exploring additional features in future research could further enhance classification performance and support direct diabetes detection. Diabetes mellitus is a chronic condition characterized by excessive glucose accumulation in the bloodstream [58]. A key complication of diabetes is autonomic nervous system (ANS) dysfunction, which significantly affects heart rate variability (HRV). Disruptions in glucose metabolism reduce parasympathetic and sympathetic activity, measurable through PPG morphology and HRV analysis [4]. In type 2 diabetes, arterial stiffness is a significant characteristic, impacting the differentiation between diabetic and non-diabetes conditions [58]. Factors such as arterial wall rigidity, viscosity, and cardiac depolarization affect the PPG waveform, suggesting diabetes may be detected through PPG signal analysis [59].

3.3 Estimation results

Linear regression for combined dataset. Figure 6 displays a linear regression graph illustrating the relationship between ADC values and instant blood glucose levels based on the analysis of the combined data (200 samples). The resulting regression equation is $Y = -0{,}5947X + 341{,}36$ with a determination coefficient of

0.8348, which indicates a strong relationship between the ADC value and invasive blood glucose level.

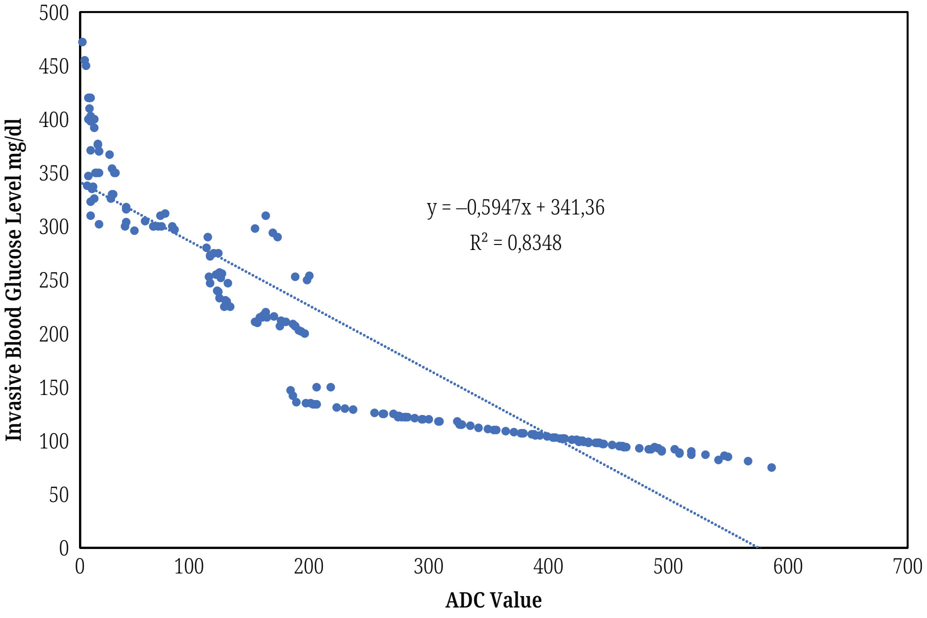


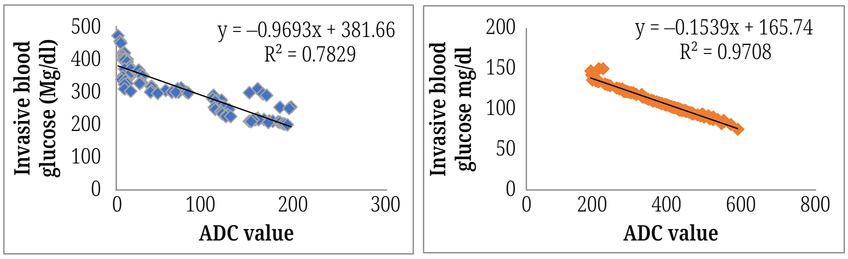
Fig. 6. Linear regression curve without stratification

The correlation coefficient between ADC values and glucose concentration is 0.91, indicating a very high correlation [42]. The coefficient of determination (R^2) reveals that the ADC value explains 83.48% of the variance in blood glucose levels. This regression model predicted non-invasive blood glucose levels on a training dataset of 200 samples and a testing dataset of 50 samples. The model's performance, evaluated using MARD, RMSE, and Pearson correlation, is summarized in Table 4.

Table 4. Estimation performance without stratification

Group	Regression	MARD (%)	RMSE (mg/dl)	Pearson Correlation
Diabetes	Simple linear regression: $y = -0.5947x + 341.36$ $R^2 = 0.8348$	9.7	40.1	0.88
Non diabetes		34.3	46.2	0.98
Overall Data		**20.3**	**39.3**	**0.92**

Linear regression for stratified modeling. The selection of diabetic and non-diabetic as stratification parameters is based on the distinct physiological and PPG signal differences between these groups. Diabetes alters arterial stiffness, blood viscosity, and tissue reflectance, resulting in significant signal variability. Combining these heterogeneous datasets into a single model increases estimation bias and reduces predictive accuracy. Stratification addresses this by creating homogeneous subsets, improving model alignment with group-specific characteristics, and enhancing accuracy. Linear regression analysis was performed separately for the diabetic and non-diabetic strata. Figure 7a illustrates the linear relationship for the diabetic stratum, while Figure 7b represents the non-diabetic stratum.



a) Diabetes stratum **b) Non-diabetes stratum**

Fig. 7. Linear regression curve with stratification

The linear regression model shows a correlation coefficient (R) of 0.88 for the diabetes stratum and 0.98 for the non-diabetes stratum. The diabetes stratum exhibits a high correlation [45], with the ADC value explaining 78.29% of the variability in invasive blood glucose levels, leaving 21.71% due to other factors [45]. In contrast, the non-diabetes stratum demonstrates a very high correlation [45], with 97.08% of blood glucose variability explained by the ADC value and 2.92% due to other factors. Unexplained variability or other factors that influence the results can arise from physiological factors of diabetic and normal subjects [3] (skin thickness variation, skin color variation, muscle thickness variation and tissue reflectance, heart rate or cardiac cycle phase, blood pressure, blood volume, aging [59], and blood circulation [60]), technical factors (such as sensor quality, sensor noise, and ADC resolution), or environmental factors (such as temperature and humidity, motion artifacts).

These findings demonstrate a strong linear relationship between ADC values and blood glucose levels, establishing ADC as a reliable predictor. The lower coefficient of determination in the diabetes stratum indicates that factors beyond PPG signal voltage significantly influence glucose levels in diabetic individuals. Stratifying the prediction models for non-diabetes and diabetes groups improves accuracy, with linear regression proving effective for the non-diabetes stratum. However, the diabetes stratum requires further model refinement. To address this, the study evaluates regression models—linear, logarithmic, polynomial, exponential, and power—within each stratum, assessing their performance using MARD, RMSE, and Pearson correlation, as summarized in Table 5.

Table 5. Estimation performance with stratification

Strata	Regression	MARD (%)	RMSE (mg/dl)	Pearson Corr
Diabetes	Linear: $y = -0.9693x + 381.66$ $R^2 = 0.7829$	8.3	31.6	0.88
	Logarithmic: $y = -52.51\ln(x) + 503.96$ $R^2 = 0.8474$	7.5	26.5	0.92
	A Second – order – Polynomial: $y = 0.0052x^2 - 1.8851x + 399.59$ $R^2 = 0.8259$	7.4	27.9	0.91
	Exponential: $y = 383.89e - 0.003x$ $R^2 = 0.8052$	8.7	30.9	0.89
	Power: $y = 566.86x - 0.171$ $R^2 = 0.8359$	8	28.0	0.91

(Continued)

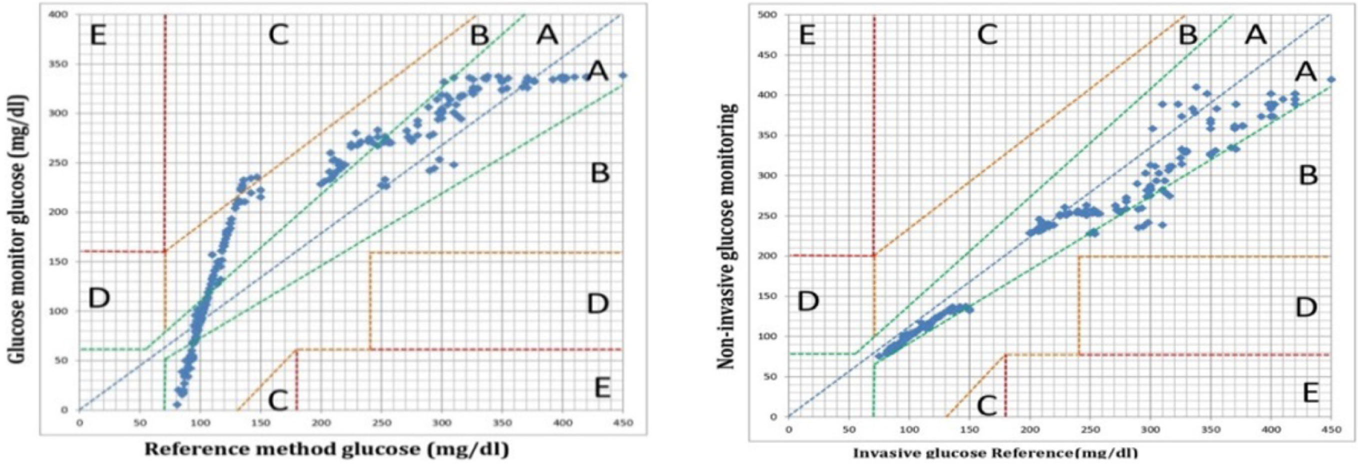
Table 5. Estimation performance with stratification *(Continued)*

Strata	Regression	MARD (%)	RMSE (mg/dl)	Pearson Corr
Non-diabetes	**Linear:** $y = -0.1522x + 164.34$ $R^2 = 0.9727$	1.4	3.13	0.98
	Logarithmic: $y = -51.93\ln(x) + 413.48$ $R^2 = 0.9723$	1.6	2.9	0.98
	A Second – order – Polynomial: $y = 0.0001x^2 - 0.2355x + 179.26$ $R^2 = 0.9764$	2.3	3.39	0.98
	Exponential: $y = 180.81e - 0.001x$ $R^2 = 0.9828$	16.6	17.5	0.985
	Power: $y = 1677x - 0.468$ $R^2 = 0.9619$	0.022	3.45	0.977
Overall data: Non-diabetes strata uses linear and diabetes strata uses logarithmic regression		**4.5**	**18.9**	**0.985**

The regression analysis demonstrates the effectiveness of stratified modelling in improving glucose estimation accuracy by tailoring regression models to the unique characteristics of diabetic and non-diabetic groups. For the diabetes stratum, logarithmic regression outperformed other models, achieving an MARD of 7.5%, an RMSE of 26.5 mg/dL, and a Pearson correlation of 0.92, compared to linear regression's MARD of 8.3%, RMSE of 31.6 mg/dL, and Pearson correlation of 0.88. These results reflect the ability of logarithmic regression to accurately capture the non-linear relationship between PPG signal voltage and glucose levels caused by physiological complexities in diabetic individuals. In contrast, the non-diabetes stratum exhibited a strong linear relationship, where linear regression achieved the best performance with an MARD of 1.4%, an RMSE of 3.13 mg/dL, and a Pearson correlation of 0.98. Although the power regression model in the non-diabetes group achieved a lower MARD (0.022%), its higher RMSE (3.45 mg/dL) suggests that linear regression offers better practicality and consistency for this population.

Logarithmic regression is particularly effective for the diabetes category because it accounts for the non-linear trends in PPG signal voltage caused by altered physiological factors such as arterial stiffness, blood viscosity, and tissue reflectance, which are significantly affected in diabetic individuals. As glucose levels increase, these factors diminish variations in PPG signal amplitude, creating a non-linear pattern that linear regression, assuming constant change, cannot adequately model. Logarithmic regression aligns more closely with the exponential-like variations in PPG signals, making it better suited for capturing the intricate dynamics of glucose metabolism and cardiovascular alterations in diabetic populations. By combining logarithmic regression for the diabetes stratum with linear regression for the non-diabetes stratum, the model achieved an overall MARD of 4.5%, RMSE of 18.9 mg/dL, and Pearson correlation of 0.985, underscoring the clear advantage of stratified modelling in reducing bias and improving predictive accuracy.

The non-invasive blood glucose measurements from the developed system are compared with those from the Accu Check glucometer, used as a reference for analysis via CEGA. The CEG analysis without stratification is shown in Figure 8a, and the stratified analysis is presented in Figure 8b.



a) Without Stratification **b) With Stratification**

Fig. 8. CEG for overall data (250 samples)

Without stratification, the CEG analysis places 26% in Region A, 71.6% in Region B, and 2.6% in Region C. There were 97.6% results in clinically acceptable areas (A and B). The 2.6% results in Region C represent clinically significant errors, indicating potentially dangerous results for treatment. With stratification, 100% of measurements were in clinically acceptable areas, with 96% in Region A and 4% in Region B. Stratification improved non-diabetes strata estimates, moving them entirely to Region A while the diabetes stratum improved to 92% in Region A and 8% in Region B. The comparison of the system's performance with and without stratification shows a significant improvement, including a 77.83% reduction in MARD, a 51.91% decrease in RMSE, a 0.065 increase in the Pearson correlation coefficient, and a CEG analysis showing 100% of test samples in areas A and B, indicating clinical acceptance. Table 6 present the comparative data.

Table 6. Estimation performance comparison in using stratification

System	MARD (%)	RMSE (mg/dl)	Pearson Correlation	CEG
Without Stratification	20.3	39.3	0.92	26% in region A, 71,6% region B, 2.6% region C
With Stratification	4.5	18.9	0.985	96% in region A, 4% in region B.

The estimation performance without stratification shows a MARD of 20.3% with a 95% confidence interval of ±1.78% and an RMSE of 39.3 mg/dL with a 95% confidence interval of±3.45%. In contrast, with stratification, the MARD is reduced to 4.5%, indicating a significant improvement in estimation accuracy, with a narrower 95% confidence interval of ±0.2%, and the RMSE decreases to 18.9 mg/dL with a 95% confidence interval of ±0.84%. This highlights the effectiveness of stratification in enhancing estimation accuracy. Table 7 presents previous studies that developed non-invasive blood glucose estimation.

Table 7. Estimation performance comparison

Authors	Signal	Method	Performance
Akkaya et al. [36]	PPG	Multiple Linear Regression	CEG: 79.17% in region A and 20.83% in region B
Nanayakkara et al. [61]	NIR	Least Squares Linear Regression	CEG: 73.3% in region A, 26.7% in region B
Zhou X et al. [26]	NIR	Random Forest	CEG: 80.35% in region A
Olakammi O et al. [23]	NIR	Linear regression	CEG: 62% in region A, 23% in region B, 15% in region C
Wei Y et al. [27]	PPG	Random Forest	MARD:12.19% CEG: 87.05% in region A.
Argiielo-Prada E.j et al. [51]	NIR PPG + US	Statistical Analysis	CEG: 91.57% in region A, 9.43% in region B.
E.-Y. Park et al. [62]	InfraRed	UOS technique	MARD: 26.6%
A.M. Joshi et al. [63]	NIRS	Multiple Polynomial Regression degree 3	MARD capillary glucose: 6.07% Serum glucose: 4.86%
A. Hina et al. [34]	PPG	GPR exponential	MARD: 8.97%
A. Hina et al. [33]	PPG	FGSVR	$R^2 = 0.937$, MARD: 7.62%
A. Hina et al. [29]	PPG	Ensemble Boosted Tree	MARD: 5.83%
M. A. Al-Dhaheri et al. [24]	PPG	Simple Linear Regression	RMSE: 10.44 mg/dl, R^2: 0.839 CEG: clinically acceptable
G. Hammour et al. [30]	PPG	Ensemble Regression	CEG: 82% in region A, 18% in region B
Proposed method	PPG	Stratified modelling (classification and simple regression)	MARD: 4.5% RMSE: 18.9 mg/dl; R: 0.985 CEG: 96% in region A, 4% in region B

As shown in Table 7, previous studies predicted blood glucose concentrations using aggregated data from diabetic and non-diabetic populations. This approach likely increased dataset heterogeneity, making it more challenging to establish an accurate predictive model. For instance, [21] applied linear regression to aggregated data, resulting in 15% of measurements falling in Region C of the CEG. In this study, without stratification, 2.6% of measurements were in Region C. After applying stratification, 100% of outcomes fell within clinically acceptable regions, with significant improvements in MARD and RMSE metrics. Stratification enhances accuracy and robustness by dividing heterogeneous data into more homogeneous subsets, aligning the model with the inherent characteristics of each group. However, the effectiveness of this approach depends on a reliable classification phase to ensure the correct predictive model is applied.

The dataset in this study comprises 200 training samples and 50 testing samples, which is sufficient to demonstrate the effectiveness of the stratified modelling approach, which may limit the model's robustness and generalizability. The relatively small and homogeneous dataset and the absence of validation on external data restrict its applicability to diverse populations. Expanding the dataset and validating with independent, more varied samples would enhance the model's reliability and provide more substantial evidence of its generalizability.

The current analysis relies on ADC voltage and age as predictors, chosen for their direct relevance to estimating glucose levels. These features sufficiently demonstrate the method's potential to prove the hypothesis that stratification can improve estimation accuracy. Expanding the feature set to include features for classification, such as stiffness index (SI), heart rate variability (HRV), and body mass index, and features for estimation, such as logarithmic energy entropy, Kaiser-Teager energy feature (KTE), spectral entropy, and signal voltages, could further improve the accuracy and robustness of the model. Future studies should explore incorporating such features to improve the generalizability of the proposed method. This study focuses on model development, but future research should explore integration into mobile health devices and clinical workflows. This would require real-time data processing, device compatibility testing, and compliance with medical standards to enable practical non-invasive glucose monitoring and improved diabetes management.

4 CONCLUSION

Implementing a stratified modeling-machine learning approach enhances the accuracy of non-invasive blood glucose concentration estimates. This study shows a 77.83% reduction in MARD and a 51.91% decrease in RMSE, with a 0.065 increase in the Pearson correlation coefficient. Additionally, stratified modeling improves CEG results, with 100% of measurements in clinically acceptable regions A and B. Stratification yields better performance in the non-diabetes stratum compared to the diabetic stratum, where a linear regression model suits the non-diabetes data and a logarithmic regression model is more effective for diabetes. However, further accuracy improvements can be achieved with more robust classification and prediction models.

5 ACKNOWLEDGMENT

This research is fully funded by the Ministry of Research and Higher Education by contract number: 041/E5/PG.02.00.PL/2024 in 2024.

6 REFERENCES

[1] Conferences and Congresses, "International diabetes federation," *The Lancet*, vol. 266, no. 6881, pp. 134–137, 1955. https://doi.org/10.1016/S0140-6736(55)92135-8

[2] P. Saeedi *et al.*, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes Atlas, 9th edition," *Diabetes Res. Clin. Pract.*, vol. 157, p. 107843, 2019. https://doi.org/10.1016/j.diabres.2019.107843

[3] E. Susana, K. Ramli, P. D. Purnamasari, and N. H. Apriantoro, "Non-invasive classification of blood glucose level based on photoplethysmography using time–frequency analysis," *Inf.*, vol. 14, no. 3, p. 145, 2023. https://doi.org/10.3390/info14030145

[4] M. S. Asekar, "Development of portable non-invasive blood glucose measuring device using NIR spectroscopy," in *2018 Second Int. Conf. Intell. Comput. Control Syst.*, no. Iciccs, 2018, pp. 572–575. https://doi.org/10.1109/ICCONS.2018.8663039

[5] D. Ghozzi, Y. Manai, and K. Nouri, "Comparison of linear and non-linear regression models for non-invasive blood glucose measurement," *J. Comput. Sci.*, vol. 15, no. 11, pp. 1607–1616, 2019. https://doi.org/10.3844/jcssp.2019.1607.1616

[6] B. Javid, F. G. Faranak, and F. S. Zakeri, "Noninvasive optical diagnostic techniques for mobile blood glucose and bilirubin monitoring," *J. Med. Signals Sens.*, vol. 8, no. 3, pp. 125–139, 2018. https://doi.org/10.4103/jmss.JMSS_8_18

[7] A. T. Priyoti, S. J. Jim, S. Hossain, S. Mahmud, S. Salvin, and A. Bhattacharjee, "Non-invasive blood glucose measurement using near infra-red spectroscopy," *IEEE Reg. 10 Humanit. Technol. Conf. R10-HTC*, 2019, pp. 1139–1142. https://doi.org/10.1109/R10-HTC47129.2019.9042473

[8] S. Rana, J. Verma, and A. K. Gautam, "Online and biomedical engineering," *Int. J. Online Biomed. Eng.*, vol. 19, no. 9, pp. 122–130, 2023. https://doi.org/10.3991/ijoe.v19i09.37065

[9] I. Ndakara Abubakar, M. Essabbar, and H. Saikouk, "Optimizing blood glucose regulation in type 1 diabetes patients via genetic algorithm-based fuzzy logic controller considering substantial meal protocol," vol. 20, no. 6, pp. 61–85, 2024. https://doi.org/10.3991/ijoe.v20i06.46929

[10] S. A. Aswathy and L. Anusree, "Non invasive blood glucose detection along with an assistive diabetes monitoring app," in *Proc. 2018 Int. Conf. Curr. Trends Towar. Converging Technol.* ICCTCT 2018, pp. 1–5, 2018. https://doi.org/10.1109/ICCTCT.2018.8551094

[11] J. Yadav, A. Rani, V. Singh, and B. Moahn, "Comparative study of different measurement sites using NIR based non-invasive glucose measurement system," *Procedia Computer Science*, vol. 70, pp. 469–475, 2015. https://doi.org/10.1016/j.procs.2015.10.082

[12] R. P. Jenie *et al.*, "Device for noninvasive optical measurement of blood glucose level based on discrete Fourier transform and fast artificial neural network," *J. Med. Devices, Trans. ASME*, vol. 13, no. 4, 2019. https://doi.org/10.1115/1.4044336

[13] K. Sarkar, D. Ahmad, S. K. Singha, and M. Ahmad, "Design and implementation of a non-invasive blood glucose monitoring device," in *2018 21st Int. Conf. Comput. Inf. Technol.* (ICCIT 2018), 2018, pp. 1–5. https://doi.org/10.1109/ICCITECHN.2018.8631942

[14] N. D. Nanayakkara, S. C. Munasingha, and G. P. Ruwanpathirana, "Non-invasive blood glucose monitoring using a hybrid technique," *MERCon 2018 – 4th Int. Multidiscip. Moratuwa Eng. Res. Conf.*, 2018, pp. 7–12. https://doi.org/10.1109/MERCon.2018.8421885

[15] J. H. Andersen *et al.*, "Bioimpedance and NIR for non-invasive assessment of blood glucose," *J. Electr. Bioimpedance*, vol. 10, no. 1, pp. 133–138, 2019. https://doi.org/10.2478/joeb-2019-0019

[16] J. Yadav, A. Rani, V. Singh, and B. M. Murari, "Levenberg – Marquardt-based non-invasive blood glucose measurement system," *IeTe Journal of Research*, vol. 64, no. 1, pp. 116–123, 2017. https://doi.org/10.1080/03772063.2017.1351313

[17] S. Habbu, M. Dale, and R. Ghongade, "Estimation of blood glucose by non-invasive method using photoplethysmography," *Sadhana – Acad. Proc. Eng. Sci.*, vol. 44, 2019. https://doi.org/10.1007/s12046-019-1118-9

[18] J. L. Moraes, M. X. Rocha, G. G. Vasconcelos, J. E. Vasconcelos Filho, V. H. C. de Albuquerque, and A. R. Alexandria, "Advances in photopletysmography signal analysis for biomedical applications," *Sensors*, vol. 18, no. 6, pp. 1–26, 2018. https://doi.org/10.3390/s18061894

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

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

[21] S. Deepthi, N. S. Valke, J. P. Parvathy, A. K. Sai, and G. Sajjan, "Non-invasive blood glucose measurement system," in *2019 5th IEEE Int. WIE Conf. Electr. Comput. Eng. WIECON-ECE 2019 – Proc.*, 2019, pp. 1–4. https://doi.org/10.1109/WIECON-ECE48653.2019.9019919

[22] S. Venkataramanan, D. Kamble, A. Bairolu, A. Singh, and R. Rao, "A novel heart rate and non-invasive glucose measuring device," in *Proc. 2017 IEEE Int. Conf. Commun. Signal Process*, ICCSP 2017, 2018, pp. 1796–1800. https://doi.org/10.1109/ICCSP.2017.8286704

[23] O. Olakanmi, I. A. Kamil, and O. P. Atilola, "NIBG: An efficient near-infrared spectroscopy-based device for tele-monitoring blood glucose level of diabetes outpatients in an e-health system," *Int. J. Biomed. Eng. Technol.*, vol. 30, no. 2, pp. 95–112, 2019. https://doi.org/10.1504/IJBET.2019.100553

[24] M. A. Al-Dhaheri, N. E. Mekkakia-Maaza, H. Mouhadjer, and A. Lakhdari, "Noninvasive blood glucose monitoring system based on near-infrared method," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 2, pp. 1736–1746, 2020. https://doi.org/10.11591/ijece.v10i2.pp1736-1746

[25] Y. Yu, J. Huang, J. Zhu, and S. Liang, "An accurate noninvasive blood glucose measurement system using portable near-infrared spectrometer and transfer learning framework," *IEEE Sens. J.*, vol. 21, no. 3, pp. 3506–3519, 2021. https://doi.org/10.1109/JSEN.2020.3025826

[26] X. Zhou, R. Lei, B. W.-K. Ling, and C. Li, "Joint empirical mode decomposition and singular spectrum analysis based pre-processing method for wearable non-invasive blood glucose esrimation," in *2019 IEEE 4th International Conference on Signal and Image Processing*, 2019, pp. 259–263. https://doi.org/10.1109/SIPROCESS.2019.8868879

[27] Y. Wei, B. W. K. Ling, Q. Liu, and J. Liu, "Multi-models with averaging in feature domain for non-invasive blood glucose estimation," in *ISPCE-ASIA 2022 – IEEE Int. Symp. Prod. Compliance Eng. – Asia 2022*, 2022, pp. 1–6. https://doi.org/10.1109/ISPCE-ASIA57917.2022.9971019

[28] Y. Sun *et al.*, "Random forest analysis of combined millimeter-wave and near-infrared sensing for non-invasive glucose detection," *IEEE Sens. J.*, vol. 23, no. 17, pp. 20294–20309, 2023. https://doi.org/10.1109/JSEN.2023.3293248

[29] A. Hina, S. Minto, and W. Saadeh, "A 208μW PPG-based glucose monitoring SoC using ensembled boosted trees," in *20th IEEE Int. Interreg. NEWCAS Conf. NEWCAS 2022 – Proc.*, 2022, pp. 241–245. https://doi.org/10.1109/NEWCAS52662.2022.9842163

[30] G. Hammour and D. P. Mandic, "An in-ear PPG-based blood glucose monitor: A proof-of-concept study," *Sensors*, vol. 23, no. 6, p. 3319, 2023. https://doi.org/10.3390/s23063319

[31] C.-T. Yen, U. Chen, G.-C. Wang, and Z.-X. Chen, "Non-invasive blood glucose estimation system based on a and bioelectrical impedance measuring," *Sensors*, vol. 22, no. 12, pp. 1–13, 2022. https://doi.org/10.3390/s22124452

[32] Q. Pham, T. Nguyen, A. H. Ba, H. Ngo, H. Vo, and N. Tran, "An embedded system for non-invasive glucose monitoring," in *2023 Int. Conf. Multimed. Anal. Pattern Recognit.*, 2023, pp. 1–6. https://doi.org/10.1109/MAPR59823.2023.10289110

[33] A. Hina and W. Saadeh, "A Noninvasive glucose monitoring SoC based on single wavelength photoplethysmography," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 3, pp. 504–515, 2020. https://doi.org/10.1109/TBCAS.2020.2979514

[34] A. Hina, H. Nadeem, and W. Saadeh, "A single LED photoplethysmography-based non-invasive glucose monitoring prototype system," in *Proc. – IEEE Int. Symp. Circuits Syst.*, 2019, pp. 1–5. https://doi.org/10.1109/ISCAS.2019.8702747

[35] P. P. Pai, P. K. Sanki, S. K. Sahoo, A. De, S. Bhattacharya, and S. Banerjee, "Cloud computing-based non-invasive glucose monitoring for diabetic care," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 65, no. 2, pp. 663–676, 2018. https://doi.org/10.1109/TCSI.2017.2724012

[36] I. Akkaya, E. Selim, M. Altintas, and M. Engin, "Power spectral density-based nearinfrared sub-band detection for noninvasive blood glucose prediction in both in-vitro and in-vivo studies," *J. Innov. Opt. Health Sci.*, vol. 11, no. 6, pp. 1–12, 2018. https://doi.org/10.1142/S1793545818500359

[37] G. Noroozi, "Data heterogeneity and its implications for fairness," *Electron. Thesis Diss. Repos.*, 2023.

[38] A. Nunes, T. Trappenberg, and M. Alda, "The definition and measurement of heterogeneity," *Transl. Psychiatry*, vol. 10, pp. 1–13, 2020. https://doi.org/10.1038/s41398-020-00986-0

[39] O. J. Ibidoja, F. P. Shan, J. Sulaiman, and M. K. M. Ali, "Detecting heterogeneity parameters and hybrid models for precision farming," *J. Big Data*, vol. 10, no. 1, 2023. https://doi.org/10.1186/s40537-023-00810-8

[40] A. Karpatne, A. Khandelwal, S. Boriah, and V. Kumar, "Predictive learning in the presence of heterogeneity and limited training data," in *SIAM Int. Conf. Data Min. 2014, SDM 2014*, 2014, pp. 253–261. https://doi.org/10.1137/1.9781611973440.29

[41] J. Tuck and S. Boyd, "Fitting Laplacian regularized stratified models," *SSRN*, 2020. https://doi.org/10.2139/ssrn.3608009

[42] D. T.-M. Dinh, V. A. Truong, A. N.-P. Tran, H. X. Le, and H. T.-T. Pham, "Non-invasive glucose monitoring system utilizing near-infrared technology," in *7th International Conference on the Development of Biomedical Engineering in Vietnam (BME7), IFMBE Proceedings 69*, Cham: Springer International Publishing, 2020, pp. 401–405. https://doi.org/10.1007/978-981-13-5859-3_71

[43] T. W. Gamessa and D. Suman, "Non-invasive blood glucose monitoring using visible laser light non-invasive blood glucose monitoring using visible laser light," *Research Journal of Pharmacy and Technology*, vol. 12, no. 2, 2019. https://doi.org/10.5958/0974-360X.2019.00144.6

[44] B. A. Al-Ghali, H. H. Alzubeidy, and Z. T. Al-Dahan, "Non-invasive measurement of arterial pH during cardiopulmonary bypass," *Int. J. online Biomed. Eng.*, vol. 18, no. 10, pp. 190–198, 2022. https://doi.org/10.3991/ijoe.v18i10.31523

[45] P. Schober and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesth. Analg.*, vol. 126, no. 5, pp. 1763–1768, 2018. https://doi.org/10.1213/ANE.0000000000002864

[46] H. Mondal and S. Mondal, "Clarke error grid analysis on graph paper and Microsoft Excel," *J. Diabetes Sci. Technol.*, vol. 14, no. 2, p. 499, 2019. https://doi.org/10.1177/1932296819890875

[47] N. A.rocha Jr and R. San Jose Reyes, "Non-invasive glucose monitoring with combinations of near infrared spectroscopy and metabolic heat conformation technology using multivariate analysis approach," in *ACM Int. Conf. Proceeding Ser.*, 2021, pp. 83–89. https://doi.org/10.1145/3431943.3431957

[48] W. V. Gonzales, A. T. Mobashsher, and A. Abbosh, "The progress of glucose monitoring—A review of invasive to minimally and non-invasive techniques, devices and sensors," *Sensors*, vol. 19, no. 4, p. 800, 2019. https://doi.org/10.3390/s19040800

[49] M. Naresh and S. Peddakrishna, "Non-invasive glucose measurement using 950 nm reflective short wave NIR technique," *Res. Biomed. Eng.*, vol. 39, pp. 747–757, 2023. https://doi.org/10.1007/s42600-023-00308-5

[50] M. Ionescu, "Glucometry and pulse oximetry-comparative noninvasive methods for determining blood glucose," in *2019 7th E-Health Bioeng. Conf. EHB 2019*, 2019, pp. 1–4. https://doi.org/10.1109/EHB47216.2019.8970001

[51] E. J. Argüello-Prada and S. M. Bolaños, "On the role of perfusion index for estimating blood glucose levels with ultrasound-assisted and conventional finger photoplethysmography in the near-infrared wavelength range," *Biomed. Signal Process. Control*, vol. 86, p. 105338, 2023. https://doi.org/10.1016/j.bspc.2023.105338

[52] V. R. Reddy, A. D. Choudhury, S. Jayaraman, N. K. Thokala, P. Deshpande, and V. Kaliaperumal, "PerDMCS: Weighted fusion of PPG signal features for robust and efficient diabetes mellitus classification," in *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (*BIOSTEC 2017*) – SmartMedDev*, 2017, pp. 553–560. https://doi.org/10.5220/0006297205530560

[53] N. Nirala, R. Periyasamy, B. K. Singh, and A. Kumar, "Detection of type-2 diabetes using characteristics of toe photoplethysmogram by applying support vector machine," *Biocybern. Biomed. Eng.*, vol. 39, no. 1, pp. 38–51, 2019. https://doi.org/10.1016/j.bbe.2018.09.007

[54] C. Hettiarachchi and C. Chitraranjan, "A machine learning approach to predict diabetes using short recorded photoplethysmography and physiological characteristics," in *Artificial Intelligence in Medicine.* AIME 2019. in Lecture Notes in Computer Science, D. Riaño, S. Wilk, and A. ten Teije, Eds., vol. 11526, Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-21642-9_41

[55] Y. K. Qawqzeh, A. S. Bajahzar, M. Jemmali, M. M. Otoom, and A. Thaljaoui, "Classification of diabetes using photoplethysmogram (PPG) waveform analysis: Logistic regression modeling," *Biomed. Res. Int.*, vol. 2020, no. 1, 2020. https://doi.org/10.1155/2020/3764653

[56] A. Prabha, J. Yadav, A. Rani, and V. Singh, "Non-invasive diabetes mellitus detection system using machine learning techniques," in *Proc. Conflu. 2021 11th Int. Conf. Cloud Comput. Data Sci. Eng.*, 2021, pp. 948–953. https://doi.org/10.1109/Confluence51648.2021.9377138

[57] J. Chu, W.-T. Yang, T.-H. Hsieh, and F.-L. Yang, "One-minute finger pulsation measurement for diabetes rapid screening with 1.3% to 13% false-negative prediction rate," *Biomed. Stat. Informatics*, vol. 6, no. 1, pp. 6–13, 2021. https://doi.org/10.11648/j.bsi.20210601.12

[58] C. Ouyang, P. Zhou, and Z. Gan, "A non-invasive test method for type-2 diabetes mellitus by pulse waveform fitting," *Biomed. Signal Process. Control*, vol. 60, p. 102000, 2020. https://doi.org/10.1016/j.bspc.2020.102000

[59] S. Zanelli, M. Ammi, M. Hallab, and M. A. El Yacoubi, "Diabetes detection and management through photoplethysmographic and electrocardiographic signals analysis: A systematic review," *Sensors*, vol. 22, no. 13, pp. 1–28, 2022. https://doi.org/10.3390/s22134890

[60] H. Tjahjadi, H. Sudaryanto, A. Budi Rahmanto, A. V. Lesmana, A. Ilham Irianto, and O. Alifian, "A review of non-invasive monitoring of blood glucose levels based on photoplethysmography signals using artificial intelligence," in *Proc. – ICACSIS 2022 14th Int. Conf. Adv. Comput. Sci. Inf. Syst.*, 2022, pp. 111–116. https://doi.org/10.1109/ICACSIS56558.2022.9923513

[61] N. D. Nanayakkara, S. C. Munasingha, and G. P. Ruwanpathirana, "Non-invasive blood glucose monitoring using a hybrid technique," *MERCon 2018 – 4th Int. Multidiscip. Moratuwa Eng. Res. Conf.*, 2018, pp. 7–12. https://doi.org/10.1109/MERCon.2018.8421885

[62] E.-Y. Park, J. Baik, H. Kim, S.-M. Park, and C. Kim, "Ultrasound-modulated optical glucose sensing using a 1645 nm laser," *Sci. Rep.*, vol. 10, 2020. https://doi.org/10.1038/s41598-020-70305-6

[63] A. M. Joshi, P. Jain, S. P. Mohanty, and N. Agrawal, "IGLU 2.0: A new wearable for accurate non-invasive continuous serum glucose measurement in IoMT framework," *IEEE Trans. Consum. Electron.*, vol. 66, no. 4, pp. 327–335, 2020. https://doi.org/10.1109/TCE.2020.3011966

7 AUTHORS

Fitrilina earned her Bachelor's degree in Electrical Engineering from Universitas Andalas, Indonesia, and her Master's degree in Electrical Engineering from Institut Teknologi Bandung, Indonesia. She is pursuing a PhD in the Electrical Engineering Department at Universitas Andalas (E-mail: fitrilina@unib.ac.id).

Muhammad Ilhamdi Rusydi (IEEE Member) earned his Bachelor's and Master's degrees in Electrical Engineering from the Institute of Technology Bandung, Indonesia, and his Ph.D. from Gifu University, Japan. Since 2005, Universitas Andalas has employed him as a Lecturer, and he was promoted to full Professor in 2023. His teaching responsibilities include biosignals, robotics, and control systems. His research focuses on control robotics and assistive technology for medical applications. Currently, he serves as the Vice Dean for Academic and Student Affairs at the Faculty of Engineering, Universitas Andalas (E-mail: rusydi@eng.unand.ac.id).

Rahmadi Kurnia received his Bachelor's (1995) and Master's (1998) degrees in Electrical Engineering from Universitas Indonesia. He earned his doctorate in 2006 from Saitama University, Japan. He is currently working as a Lecturer in the Electrical Engineering Department at Universitas Andalas (E-mail: kurnia@eng.unand.ac.id).

Noverika Windasari completed her general medical education at the Faculty of Medicine, Universitas Andalas (2010). She continued specialist education in Forensic Medicine and Medicolegal Studies at the Faculty of Medicine, Pajdjadjaran University, Bandung (2016–2020). Currently, Dr. Winda is the Head of the Department of Forensics and Medicolegal, Faculty of Medicine, Universitas Andalas. She is also the Secretary of the Research Ethics Commission of the Faculty of Medicine, Universitas Andalas.

Raffi Zahrandhika Putra graduated with a Bachelor's degree in Electrical Engineering from Universitas Bengkulu, Indonesia. He is currently working as a Specialist Consultant in Mechanical and Electrical (Instrumentation Control) at a Consulting Services Company in South Jakarta, Jakarta, Indonesia (E-mail: raffizahrandhikap@gmail.com).