www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 17(2), 2026, 426-436. ISSN: 2007-1558. https://doi.org/10.61467/2007.1558.2026.v17i2.1110

Design and Evaluation of a Third-Order Regression Model for Estimating Glucose Levels from ECG Signal Features

Oscar I. Coronado-Reyes [1], Adriana del C. Téllez-Anguiano [1], Eilen García-Rodríguez [1], Luis A. Castro-Pimentel [1], Juan Carlos Olivares-Rojas
[1] TecNM / Instituto Tecnológico de Morelia, Morelia, Michoacán, México
E-mail: {oscar.cr, adriana.ta, eilen.gr, luis.cp, juan.or} @morelia.tecnm.mx

Abstract. Diabetes mellitus (DM), also known simply as diabetes, is a chronic disease that affects the absorption of glucose from food, increasing its levels in the blood, whose routine monitoring requires invasive methods. This study presents the statistical and clinical evaluation of a non-invasive Blood Glucose Level (BGL) estimation approach using a third-order multivariate regression model. A total of 271 healthy and diabetic participants aged 18–70 was analyzed. Initial Pearson correlation analysis between BGL and electrocardiographic features, Heart Rate (HR), Heart Rate Variability (HRV), R-wave, S-wave, and T-wave amplitudes, and QT interval revealed the strongest correlations for HR (0.3259) and HRV (−0.3725). Combining HR/HRV further improved the correlation to 0.5861. Model validation via ANOVA yielded a result of 0.7236. Clinical validation using Clarke's Error Grid Analysis showed 58.54% of validation data in Zone A, 39.04% in Zone B, 1.21% in Zone C, and 1.21% in Zone D, with an overall correlation of 0.8525. These findings demonstrate the potential of ECG-derived parameters as a viable non-invasive tool for BGL estimation.
Keywords: Diabetes, regression, Anova, Clarke error grid.

Article Info
Received April 2, 2025
Accepted Dec 11, 2025

1 Introduction

Diabetes mellitus (DM), commonly referred to as diabetes, is one of the leading causes of death worldwide. It is a chronic condition with no cure, characterized by elevated blood glucose levels (BGL). The human body breaks down most of the ingested food into sugar (glucose). When blood glucose levels increase, a signal is sent to the pancreas to release insulin. Insulin functions as a key, allowing glucose in the blood to enter cells where it is used for energy (*Diabetes Basics | Diabetes | CDC*, n.d.).

When there is insufficiency insulin, or when cells become resistant to its action, blood glucose levels increase. DM is a chronic condition with no cure, but if blood glucose levels are not properly managed, it can lead to severe health complications, including cardiovascular disease, vision impairment, kidney disease, and others (*Diabetes Basics | Diabetes | CDC*, n.d.).

Current global statistics on diabetes mellitus (DM) indicate that in 2021, approximately 537 million adults (aged 20–79 years) were living with DM, with nearly half unaware of their condition. It is projected that the number of individuals with DM will increase to 643 million by 2030 and 783 million by 2045 (*Datos y Cifras Sobre La Diabetes | Federación Internacional de Diabetes*, n.d.).

There are three types of diabetes mellitus (DM): type 1 DM (DM1), type 2 DM (DM2), and gestational DM (GDM). In individuals with DM1, the pancreas either does not produce insulin or produces very little. DM1 typically manifests in children and young adults. Individuals with DM1 require insulin injections to control blood glucose levels. The main symptoms of DM1 include abnormal thirst and dry mouth, unintentional weight loss, frequent urination, lack of energy, fatigue, constant hunger, blurred vision, and nocturnal enuresis (*Diabetes de Tipo 1 | Federación Internacional de Diabetes*, n.d.).

DM2 occurs when there is insulin resistance or insulin cannot function properly, then blood glucose levels continue to rise, prompting the release of more insulin. Unfortunately, for some individuals with DM2, this process can eventually exhaust the pancreas. As a result, the body produces progressively less insulin, leading to even higher blood glucose levels (hyperglycemia). DM2 accounts for approximately 90% of all cases of DM. The main symptoms of DM2 include excessive thirst and dry mouth, frequent urination, lack of energy, fatigue, delayed wound healing, recurrent skin infections, blurred vision, and tingling or numbness in the hands and feet, among others. In most cases, the disease can be prevented. The primary risk factors for DM2 are family history of DM, overweight or obesity, unhealthy diet, physical inactivity, increasing age, high blood pressure, impaired glucose tolerance (IGT), gestational diabetes, among others (*Diabetes de Tipo 2 | Federación Internacional de Diabetes*, n.d.).

The continuous monitoring and control of BGL can improve the quality of life for individuals and help prevent future complications such as vision problems, gastrointestinal issues, and, primarily, cardiovascular diseases. Regular assessment of BGL provides valuable information for managing DM, including monitoring the effects of diabetes medications, identifying whether BGL is too high or low, adjusting treatment as necessary to control glucose levels, tracking progress towards treatment goals, evaluating the impact of diet and exercise on BGL, and understanding the effects of illness or stress on BGL (Elsayed et al., 2023).

It is recommended that individuals with DM1 perform continuous blood glucose (BGL) measurements between 4 and 10 times per day or utilize continuous glucose monitoring (CGM) systems. These measurements should be taken both pre- and postprandially, as well as before and after the administration of medications or insulin injections. The frequency of BGL measurements for individuals with DM2 depends on the type and amount of insulin they use; typically, measurements are recommended before meals and at bedtime (Elsayed et al., 2023).

The traditional blood glucose meter (BGM) involves fingerstick punctures to extract a capillary blood sample for glucose analysis. These methods have technical limitations for CGM and can cause discomfort to patients (Galindo et al., 2020). However, advances in technology in recent decades have led to the development of CGM devices, which have become one of the key innovations for DM management. CGM devices offer a more user-friendly and less invasive solution. For individuals with diabetes who require insulin for disease control, CGM devices can improve quality of life by providing real-time data that assist in making informed decisions regarding diet, insulin dosage, and overall lifestyle (*Control Continuo de La Glucosa (MCG) - Federación Internacional de Diabetes*, n.d.).

CGM systems function through small sensors that are placed under the skin. These sensors are typically positioned in adipose tissue, which can securely hold the sensor in place, ensuring comfort during use. The most recommended placement sites include the posterior part of the arm, the abdomen, and the outer thigh. The sensors are connected via a needle or catheter, which transmits GBL to a receiver or mobile application. One of the advantages of using a CGM system is the ability to track glucose levels continuously, providing a comprehensive view of blood glucose fluctuations. CGM devices also help reduce the risk of both hypoglycemia (low blood glucose) and hyperglycemia (high blood glucose) by providing real-time alerts when glucose levels fall outside of range (*Control Continuo de La Glucosa (MCG) - Federación Internacional de Diabetes*, n.d.).

CGM devices also help make glucose monitoring smoother and easier by eliminating the need for frequent finger pricks throughout the day. Unlike blood glucose meters, CGMs measure glucose levels from the interstitial fluid, which is the fluid surrounding cells. Unlike finger-prick tests, which provide immediate results, there is a five- to 15-minute delay with CGM in reporting changes in glucose levels. This can cause differences in reading and how a person feels. Other drawbacks of these devices include issues with sensor adhesion, particularly during physical activity or outdoor use, connectivity problems, and software malfunctions. There are also compatibility requirements with specific smartphones or the need for additional equipment for data synchronization. Furthermore, inaccurate readings can lead to potential misinterpretations and incorrect treatment decisions (*Control Continuo de La Glucosa (MCG) - Federación Internacional de Diabetes*, n.d.).

Recent research has focused on developing techniques for non-invasive BGL measurement, with some of these methods based on near-infrared (NIR) spectroscopy, optical methods, and Raman spectroscopy (Alsunaidi et al., 2021; Biomed Phys Eng et al., n.d.; Jain et al., 2024; Luis Alfredo Castro Pimentel, Adriana del Carmen Téllez Anguiano & Kevin Raúl Hernández Franco., 2019; Sun, 2022; Zanelli et al., 2022). These studies have shown promising results for non-invasive CGM. The main problem with these techniques is that there is a great variety of physical and functional variables that intervene in the process such as: fat percentage, temperature, humidity, all skin, body hair, etc., which generates a complex implementation model for portable systems and serious errors potentially harmful to health. On the other hand, one of the main limitations of the techniques is their measurement range for blood glucose levels (BGL). These methods have shown good performance in individuals without diabetes and within the normal glucose range (100–150 mg/dL), but they tend to fail during periods of hyperglycemia or hypoglycemia.

Finally, research has been conducted to analyze the relationship between cardiac activity as recorded in an electrocardiogram (ECG), and blood glucose levels (BGL). These studies are potential alternatives for detecting periods of hyperglycemia (Aggarwal et al., 2021; Cordeiro et al., 2021; Elvebakk et al., 2019; Li et al., 2021; Swapna et al., 2020; Swapna, Soman, et al., 2018; Swapna, Vinayakumar, et al., 2018), as the function of the cardiovascular system and changes in the sympathetic and parasympathetic nervous systems are closely related to fluctuations in glucose levels (Effects of Glucagon as a Neurohormone on the Central Nervous System and Glucose Homeostasis, n.d.; Kajisa et al., 2024; Townsend, 2024; Zsombok et al., 2024).

The following study analyzes the feasibility of designing a model for estimating BGL based on the characteristics of ECG signal. For the evaluation of the proposed model, ANOVA analysis and Pearson correlation are used. This initial research provides valuable information for the future development of a non-invasive blood glucose measurement tool capable of covering a wide range of glucose levels.

2 Methodology

2.1 Participants

The volunteers had ECG recordings using the VII lead, while blood glucose levels (BGL) were measured with the commercial Accu-Chek Performa® device. The study's volunteer participants were randomly selected within an age range of 18 to 70 years (healthy individuals and those with DM2. It is important to note that the study followed the criteria established in the Declaration of Helsinki. The exclusion criteria were as follows: individuals with cardiovascular problems or diagnosed with heart disease. The study was approved by the ethics committee of the Technological Institute of Morelia.

2.2 Sampling

2.2.1 ECG signals

The electrocardiogram (ECG) is a time-amplitude recording of the heart's electrical activity, generated by the depolarization and repolarization of the atria and ventricles. This process is responsible for carrying oxygen-enriched blood to the entire human body. The ECG contains three main components: the P wave, the QRS complex, and the T wave. These waves contain crucial information, such as amplitude and duration, which are important for diagnosing cardiovascular diseases.
ECG measurement was performed using the ADS1293EVM evaluation platform which has the following specifications: 24-bit resolution, 3- to 5-lead configuration, sampling frequency setting in the range of 200 to 1900 Hz, patient isolation and protection, noise removal filtering, import of records for post-processing, free software for ECG signal acquisition. The configuration implemented for sampling is shown in Fig. 1 (A), corresponding to three leads.

2.2.2 Measurement of Blood Glucose Concentration

BGL was measured using the commercial device Accu-Chek Performa®, which has been validated according to the international standard ISO 15197:2013 (Fig. 1 (B)).

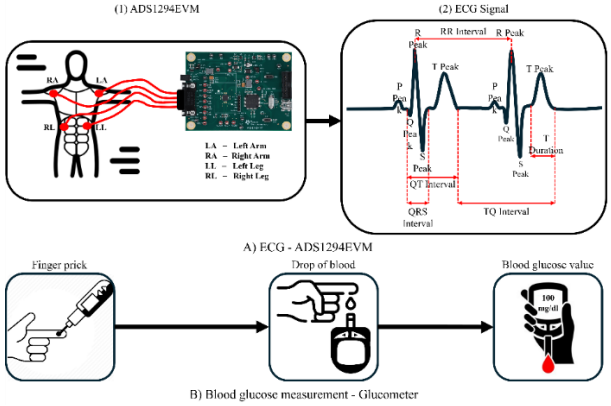

Fig. 1. Protocols for ECG and BGL Measurement.

2.3 Detection of Features in ECG Signals

Signal processing and feature detection in ECG signals were performed using MATLAB software, The results were supervised and validated by expert physicians (cardiologists). The signal processing begins with the removal of line noise (60 Hz). To achieve this, a notch filter with a central frequency of 60 Hz and a selectivity of -15 dB is designed. Subsequently, the baseline correction is performed by eliminating frequency components less than 0.5 Hz. For this, a third-order low-pass Butterworth filter with a cutoff frequency of 0.5 Hz is applied.

Once the unwanted frequency components are removed, the resulting ECG signal is plotted. The features of interests such as the R peak, the T peak, the Q peak, the onset of the Q wave, and end of the T wave—are manually identified and marked visually on the ECG signal (an 8-second recording) storing the time and amplitude value where the events occur.

The heart rate (HR) is calculated by counting the number of R peaks identified in the 8-second ECG recording. The heart rate variability (HRV) is determined by calculating the standard deviation of the RR intervals (the time between successive R waves) found in the same 8-second recording. The T peak corresponds to the average of all the T wave amplitudes detected in the signal. Similarly, the S peak is the average of all the S wave amplitudes identified in the recording. Finally, the QT interval corresponds to the average between the beginning of the Q wave and the end of the T wave of each of the cardiac cycles in the recording. The mentioned characteristics were acquired simultaneously with the GBL.

2.4 Statistical analysis

Statistical analysis is used in medicine to identify relationships or describe data and make inferences. It helps answer key questions about the data, such as testing hypotheses, describing associations (correlation), modeling relationships (regression), and measuring the effect between analyzed variables (Kanti V. Mardia et al., 2024).

The main objective of statistical analysis is to understand or determine the relationship between a response variable (output) and the explanatory variables (inputs) of a phenomenon. One of the primary techniques for describing the behavior of such phenomena is regression analysis, which models their relationship. The type of regression (linear, quadratic, cubic, etc.) depends on the distribution of the dependent variable Y. In general, the variables should be continuous and approximately normally distributed (Kanti V. Mardia et al., 2024).

2.4.1 Multivariable regression

Multiple regression (or multivariable regression) describes the relationship between a dependent variable (output) and multiple independent variables (inputs). These independent variables must be continuous and measured on a fixed interval scale.

The basic first-order regression model is given by (1):

$$Y = a + b_1x_1 + b_2x_2 + \cdots + b_nx_n + e \tag{1}$$

Where Y is the dependent variable or output, $a, b_1, b_2, \ldots, b_n$ are known parameters referred to as weights. $x_1, x_2, \ldots, x_n$ are predictor variables or inputs. Expanding (1) for each observation is obtained (2):

$$\begin{aligned} y_1 &= b_1 + b_2x_{21} + b_3x_{31} + \cdots + b_nx_{n1} + e \\ y_2 &= b_2 + b_2x_{22} + b_3x_{32} + \cdots + b_nx_{n2} + e \\ &\vdots \\ y_k &= b_k + b_2x_{2k} + b_3x_{3k} + \cdots + b_nx_{nk} + e \end{aligned} \tag{2}$$

The system of equations can be expressed in matrix form as (3):

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{21} & x_{31} & \cdots & x_{n1} \\ 1 & x_{22} & x_{32} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{2k} & x_{3k} & \cdots & x_{nk} \end{pmatrix} x \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \tag{3}$$

The procedure to start with the regression begins with the design of the matrix of predictor variables (inputs) and the calculation of weights (4).

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = (X^T X)^{-1} X^T Y \tag{4}$$

Where X is the matrix of inputs or predictor variables, Y is the output value taken in the different observations, b are the weights that are multiplied by the input variables producing an estimated result.

The general model of a multivariate linear regression is defined by (5):

$$\mathrm{Y} = \mathrm{Xb} + \mathrm{e} \tag{5}$$

Where X is the matrix of estimators, Y is the output value, b is the weight vector and e is the error in the estimation.

The quadratic or second-order regression is defined by (6):

$$\mathrm{Y} = \mathrm{b_0} + \sum_{\mathrm{i=1}}^{\mathrm{D}} \mathrm{b_i x_i} + \sum_{\mathrm{i=1}}^{\mathrm{D}} \sum_{\mathrm{j=1}}^{\mathrm{D}} \mathrm{b_{ij} x_i x_j} \tag{6}$$

Where Y is the estimated output value, $\mathrm{b_0}$, $\mathrm{b_i}$, $\mathrm{b_{ij}}$ are the parameters known as weights, $\mathrm{x_i}$ and $\mathrm{x_j}$ are the predictor variables.

The matrix model of a quadratic regression is defined by (7):

$$\mathrm{Y} = \mathrm{X^2 a} + \mathrm{Xb} + \mathrm{e} \tag{7}$$

Finally, a multivariable cubic regression or third-order regression is given by (8):

$$\mathrm{Y} = \mathrm{b_0} + \sum_{\mathrm{i=1}}^{\mathrm{D}} \mathrm{b_i x_i} + \sum_{\mathrm{i=1}}^{\mathrm{D}} \sum_{\mathrm{j=1}}^{\mathrm{D}} \mathrm{b_{ij} x_i x_j} + \cdots + \sum_{\mathrm{i=1}}^{\mathrm{D}} \sum_{\mathrm{j=1}}^{\mathrm{D}} \sum_{\mathrm{k=1}}^{\mathrm{D}} \mathrm{b_{ijk} x_i x_j x_k} \tag{8}$$

Where Y is the estimated output value, $\mathrm{b_0}$, $\mathrm{b_i}$, $\mathrm{b_{ij}}$, $\mathrm{b_{ijk}}$ are the parameters known as weights, and $\mathrm{x_i}$, $\mathrm{x_j}$, $\mathrm{x_k}$, are the predictor variables.

The matrix model described for a third-order or cubic regression is defined by (9):

$$\mathrm{Y} = \mathrm{X^3 a} + \mathrm{X^2 b} + \mathrm{Xb} + \mathrm{e} \tag{9}$$

2.4.2 Analysis of Variance (ANOVA)

ANOVA is a statistical test used to detect differences in the means of groups when there is a parametric dependent variable and one or more independent variables. ANOVA is suitable for experimental designs with a dependent variable that is a measure of the continuous parametric numerical outcome from multiple experiments across one or more independent variables (Okoye & Hosseini, 2024).

ANOVA is mathematically based on quantifying the relationship between the dependent variable and the independent variable(s). To apply an ANOVA analysis, the experimental data must meet the following criteria: they must be parametric data and follow a normal distribution (Okoye & Hosseini, 2024). ANOVA is based on three main calculations: the sum of squares, degrees of freedom, and mean square. The Sum of Squares for Regression (SSR) is defined by (10):

$$\mathrm{SSR} = \sum_{\mathrm{i=1}}^{\mathrm{n}} (\hat{y}_\mathrm{i} - \bar{\mathrm{y}})^2 \tag{10}$$

Where $\widehat{y_i}$ is the estimated output for each sample, and $\bar{y}$ is the meaning of the model's estimates.

The Sum of Squares for Error (SSE) is defined by (11):

$$SSE = \sum_{i=1}^{n} (y_i - \widehat{y_i})^2 \quad (\mathbf{11})$$

Where y_i s the actual reference value and, $\widehat{y_i}$ is the estimated output for each sample.

The Total Sum of Squares (SST), which is equivalent to the SSR and SSE, is defined by (12):

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad (\mathbf{12})$$

Where y_i is the actual reference value, and $\bar{y}$ is the meaning of the model's estimates.

The Degrees of Freedom (DF) corresponds to the number of observations, dependent variables, and independent variables used in the analysis.

The Mean Square for Regression (MSR) is defined by (13):

$$MSR = \frac{SSR}{Vi} \quad (\mathbf{13})$$

Where Vi is the number of independent variables.

The Mean Square for Error (MSE) is given by (14):

$$MSE = \frac{SSE}{n - V} \quad (\mathbf{14})$$

Where n is the number of observations and is the number of dependent and independent variables.

Finally, the Total Mean Square (MST) is defined by (15):

$$MST = \frac{SST}{n - Vd} \quad (\mathbf{15})$$

Where Vd is the number of independent variables.

The F is given by (16):

$$F = \frac{MSR}{MSE} \quad (\mathbf{16})$$

To calculate the coefficient of determination (R^2) it is defined as (17):

$$R^2 = \frac{SSR}{SST} \quad (\mathbf{17})$$

Finally, the adjusted coefficient of determination ($R^2_{ajustado}$) is defined by (18):

$$R^2_{ajustado} = 1 - \frac{MSE}{MST} \quad (\mathbf{18})$$

2.5 Validation

The Clarke error grid (Fig, 2(C)) method is a clinical technique used to evaluate the clinical significance of differences between test glucose measurement techniques and reference blood glucose measurements. The method utilizes a Cartesian plot, where the values estimated by the test technique are shown on the Y-axis, while the values obtained from the reference method are displayed on the X-axis. Zone A (clinically acceptable) corresponds to or represents glucose values that deviate from reference values by ± 20% or are in a hypoglycemic range, the ranges in this zone are clinically acceptable. Zone B (benign errors) are found above and below zone A, this zone represents those values that deviate from reference values by more than 20%. Values within zones A and B are clinically acceptable, while values included in zones C and E are potentially dangerous and there is the possibility of making clinically significant errors (*Clarke Error Grid Analysis - File Exchange - MATLAB Central*, n.d.).

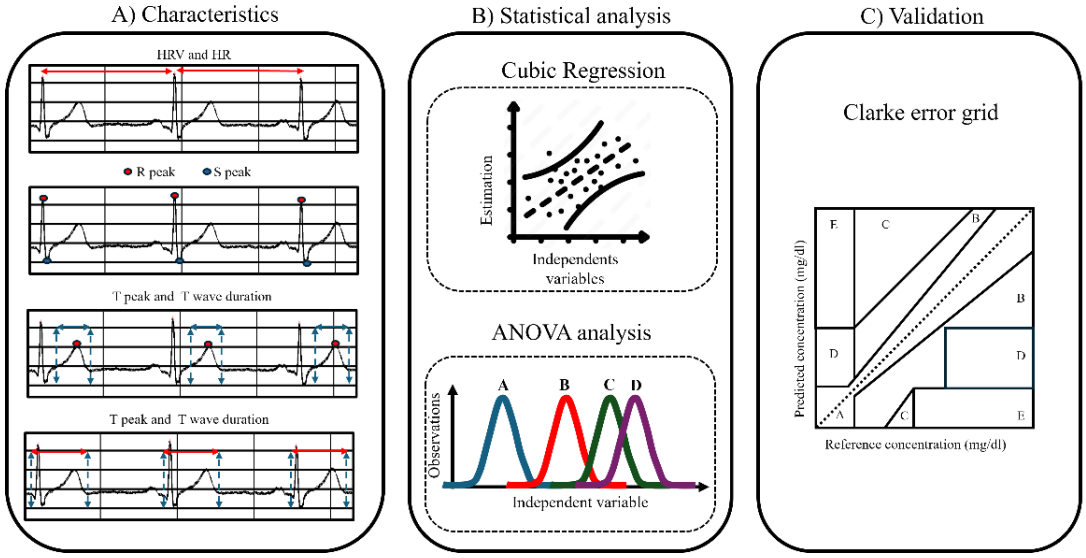

Fig. 2. Methodology implemented in the research.

3 Results

A total of 271 samples were analyzed, considering an age range of 18 to 70 years, with DM problems and high blood glucose levels. Study subjects who had heart problems were excluded; the glucose range was 65 to 500 mg/dl.

The analysis performed begins with the Pearson correlation index calculated between the predictor variables (inputs) and the output variable (glucose).

Table 1. Pearson correlation between input variables and blood glucose levels.

	HR	HRV	HRV/HR	R peak	T peak	S peak	QT interval
Glucose	0.3259	-0.3725	0.5861	0.0753	-0.1277	0.0603	-0.1155
Correlation	moderate	moderate	Stronge	Weak	Weak	None	Wak

As shown in Table 1, the highest correlation results were observed between GBL and HR as well as HRV, both of which exhibited a moderate correlation. On the other hand, the correlation with the T peak and the QT interval was found to be weak. Finally, the correlations between blood glucose levels and the R peak and S peak amplitudes were very low.

It is important to mention that the ECG signal features listed in Table 1 were used as inputs to the regression model because they have been identified as the most significant features in previous research studies.

Considering the characteristics of the ECG signal (inputs), the regression model was designed. In this case, a third-order regression was proposed due to the moderate and low correlation between the ECG features and blood glucose levels. For the design of the regression model, 70% of the dataset (190 samples) was randomly selected. In Table 2, the weights generated in the third-order multivariable equation are shown.

Table 2. The weights of the multivariable cubic regression model.

	weights		weights
1	1237.006	12	-1302.753
2	-4963.4426	13	21.509
3	0.045784	14	-2144.068
4	119201.853	15	-11.0479
5	53397.5762	16	9158.948
6	-3675.1665	17	-9.13087e-08
7	0.01274	18	9.1483
8	28.28634	19	19.3735
9	-0.000109	20	19.5802
10	-1747.109	21	-0.06467
11	-956.4725	22	2542.0438

Once the model was designed, the results were evaluated using ANOVA analysis, where the results shown in Table 3 were obtained.

Table 3. ANOVA analysis.

	SS	DF	MS
Regression	437892.533	6	62556.07614
Error	159943.4423	183	874.0078812
Total	597835.9752	189	3163.134123

From the results in Table 3, the following parameters were obtained. $f = 83.50$, the value of $R^2 = 0.7324$ and the $R^2_{adjusted} = 0.7236$. Given this information from the ANOVA analysis, it can be concluded that the $R^2_{ajustado}$ indicates that the cubic regression model is able to estimate approximately 70% of the blood glucose levels, considering the ECG signal characteristics previously mentioned.

After completing the statistical analysis of the cubic regression model, the next step was to evaluate the model using the remaining 30% of the data through the Clarke error grid method (Fig. 3).

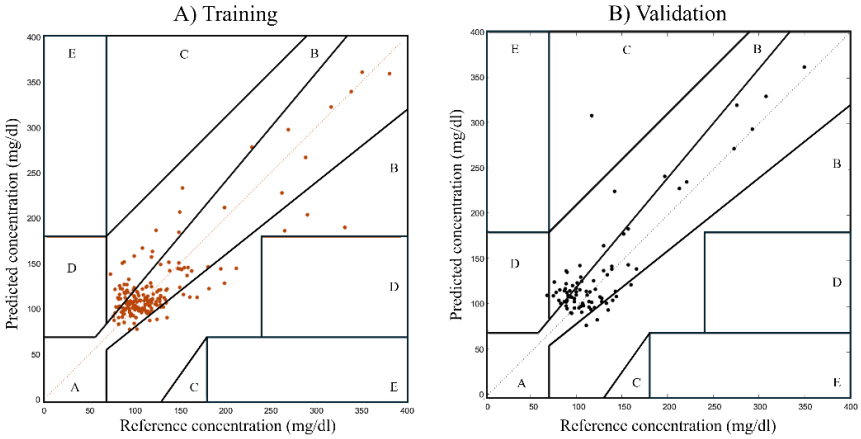

Fig. 3. Clarke error grid with the results from the evaluation of the third-order regression model using the training and validation data.

As shown in Fig. 3, the greatest dispersion of data is observed in the normal glucose range (100 – 150 mg/dl). This is because the electrocardiographic characteristics in this range present greater variation; at elevated glucose levels, the ECG characteristics show more variability, which leads to better estimation accuracy. On the other hand, the results of the estimation can be visually observed in zones A and B using the training data (Fig. 3(A)). In contrast, when analyzing the validation data (Fig. 4(B)), the estimated points are distributed across zones A, B, C, and D.

In Table 4, the training data show that 70% of the data points fall within Zone A, and 30% fall within Zone B, with a correlation value of 0.857. This correlation indicates the difference or error between the estimated and the actual (reference) values. A high correlation value suggests that the model is generally accurate in predicting glucose levels for the training data. On the other hand, the validation data showed the following distribution: 58.54% of the data points were in Zone A, 39.04% in Zone B, and 1.21% of the data points were in Zone C, with a similar percentage in Zone D. The corresponding data in zone C belongs to data from a person with DM with normal glucose levels, their ECG characteristics showed a behavior like high glucose levels due to their condition. Similarly, the data in zone D corresponds to data with blood glucose levels of 65 mg/dl, corresponding to a hypoglycemic period; in the database, by having only one data in this range, the model fails in the estimation.

Table 4. shows the summary of the estimations localized within the zones of the Clarke error grid.

	A		B		C		D		E		Correlation
Training data	133	70%	57	30%	-	--	-	--	-	-	0.8571
Validation data	48	58.54%	33	39.04%	1	1.21%	1	1.21%	-	-	0.8525

Table 5 shows the comparison with previous research focused on estimating blood glucose using non-invasive techniques.

Research	Measurement technique	Samples	Estimation technique	Statistical validation	Clinical validation
(Naresh et al., 2024)	Near infrared	115	Feed forward neural network	R^2 (.99)	---
(Agrawal et al., 2022)	iGLU device	99	Random forest	Accuracy (84 %)	Clarke error grid (100 % A y B)
(Agrawal et al., 2022)	Near infrared	74	Multiple Polynomial Regression	Average Error (4.8 %)	Clarke error grid (100 % A y B)
(Jain et al., 2019)	Near infrared	43	Neural network	Average Error (5.14 %)	Clarke error grid (100 % A y B)
Proposed Work	ECG	271	Multiple Polynomial Regression	R^2 (.7324)	Clarke error grid (97.6 % A y B)

The comparative results shown in Table 5 reflect the results obtained in previous research for noninvasive blood glucose estimation. The noninvasive techniques presented in Table 5 correspond to near-infrared techniques. No research has focused on estimating glucose levels using ECG recordings of cardiac activity. Therefore, compared to the near-infrared technique, the use of ECG to estimate blood glucose levels, in comparison to the near-infrared technique, shows promising results with both clinically and statistically acceptable validation.

4 Conclusions

Based on the results obtained, it can be mentioned that, with a database of 271 samples, a cubic regression model can be obtained with promising results for estimating BGL in wide ranges, taking as main characteristics the variables of HR, HRV, QT interval, amplitude of R, S and T in an 8-second ECG record. The ANOVA analysis yielded an adjusted $R^2_{adjusted}$ value of 0.7236, which indicates that statistically, the ECG characteristics can describe blood glucose levels with an accuracy of over 70.

The ECG signal features most strongly associated with BGL were HR (0.3259) and HRV (-0.3725), showing moderate correlation. When combining HR and HRV into a common factor, the correlation with BGL increased to 0.5861, reaching strong correlation levels. Although other features showed weaker correlations, they were still included in the multivariate regression model as they contributed to blood glucose level estimation, improving the R^2 value. This occurs because Pearson correlation only measures linear relationships, while regression analysis captures the combined effect of independent variables (inputs) on the dependent variable (output).

Clarke Error Grid Analysis (EGA) results demonstrated promising performance, with 58.54% of estimations in Zone A and 39.04% in Zone B. Most Zone B estimates corresponded to normal blood glucose levels (>100 mg/dL and <150 mg/dL). his is because, in healthy individuals, the ECG signal characteristics show less variation than in individuals with elevated blood glucose levels or diabetes mellitus (DM).

The critical clinical errors located in Zones C (1.21%) and D (1.21%) are primarily due to the lack of samples with BGLs below 70 mg/dL. Without records in this range, the regression model cannot estimate glucose levels in this range. On the other hand, the database contains data from people with DM who control their blood glucose levels (< 150 mg/dl), so the estimate in these samples fails due to the similarity between the characteristics of people with DM and those with blood glucose levels greater than 150 mg/dl.

Blood glucose estimation results can be improved by increasing the database, specifically in glucose ranges below 70 mg/dL. It is also possible to analyze the effect on the model of data from people with DM who monitor their BGL.

Finally, it can be mentioned that to improve the results of the estimate it is necessary to increase the sample amount in the hypoglycemic period and in people with DM who control their BGL. It is important to mention that, to reduce errors in ECG signal recordings, specific protocols must be followed when taking the ECG measurement. This protocol includes resting for 5 minutes to stabilize the heart rate and avoid any sudden movement of the electrodes or limbs that could interfere with the ECG signal. By ensuring stable conditions during the measurement process, the accuracy and reliability of the recorded ECG data can be significantly improved, leading to more precise estimations of BGL.

A multivariate cubic regression model can be easily implemented in systems such as microcontrollers, compared to other machine learning techniques, to design a non-invasive device for measuring blood glucose levels across a wide range, in conjunction with the ADS1293EVM acquisition system.

It is important to mention that ECG signal characteristics with low correlation are considered in the multivariate regression model, because these characteristics influence blood glucose levels, increasing the R^2 value. This is because the Pearson correlation only indicates the linear relationship but not the effect of the independent variables (inputs) on the dependent variable (output).

The lack of techniques to validate systems for noninvasive blood glucose measurement is a challenge for the design of estimation models. Therefore, in the work presented, statistical techniques were used to validate the estimation models and traditional techniques to evaluate commercial invasive systems (Clarke error grid). The reference values depend on commercial devices in the case of the Accu-chek Performa® device validated by international standards (ISO 15197:2013).

References

Aggarwal, Y., Das, J., Mazumder, P. M., Kumar, R., & Sinha, R. K. (2021). Heart rate variability time domain features in automated prediction of diabetes in rat. *Physical and Engineering Sciences in Medicine, 44*(1), 45–52. https://doi.org/10.1007/s13246-020-00950-8

Agrawal, H., Jain, P., & Joshi, A. M. (2022). Machine learning models for non-invasive glucose measurement: Towards diabetes management in smart healthcare. *Health and Technology, 12*(5), 955–970. https://doi.org/10.1007/s12553-022-00690-7

Alsunaidi, B., Althobaiti, M., Tamal, M., Albaker, W., & Al-Naib, I. (2021). A review of non-invasive optical systems for continuous blood glucose monitoring. *Sensors, 21*(20), 6820. https://doi.org/10.3390/s21206820

Centers for Disease Control and Prevention. (n.d.). Diabetes basics. https://www.cdc.gov/diabetes/about/index.html

Clarke error grid analysis. (n.d.). *MATLAB Central File Exchange*. https://la.mathworks.com/matlabcentral/fileexchange/20545-clarke-error-grid-analysis

Drucker, D. J. (2018). Mechanisms of action and therapeutic application of glucagon. *Cell Metabolism*. https://doi.org/10.1016/j.cmet.2018.03.001

Elsayed, N. A., Aleppo, G., Aroda, V. R., Bannuru, R. R., Brown, F. M., Bruemmer, D., Collins, B. S., Hilliard, M. E., Isaacs, D., Johnson, E. L., Kahan, S., Khunti, K., Leon, J., Lyons, S. K., Perry, M. L., Prahalad, P., Pratley, R. E., Seley, J. J., Stanton, R. C., & Gabbay, R. A. (2023). Diabetes technology: Standards of care in diabetes—2023. *Diabetes Care, 46*(Suppl 1), S111–S127. https://doi.org/10.2337/DC23-S007

Elvebakk, O., Tronstad, C., Birkeland, K., & Jenssen, T. (2020). A multiparameter model for non-invasive detection of hypoglycemia. *Biomedical Physics & Engineering Express*. https://doi.org/10.1088/2057-1976/abe778

Galindo, R. J., Aleppo, G., Klonoff, D. C., Spanakis, E. K., Agarwal, S., Vellanki, P., Olson, D. E., Umpierrez, G. E., Davis, G. M., & Pasquel, F. J. (2020). Implementation of continuous glucose monitoring in the hospital. *Journal of Diabetes Science and Technology, 14*(4), 822–832. https://doi.org/10.1177/1932296820932903

International Diabetes Federation. (n.d.). Continuous glucose monitoring. https://idf.org/es/about-diabetes/continuous-glucose-monitoring/

International Diabetes Federation. (n.d.). Diabetes facts and figures. https://idf.org/es/about-diabetes/diabetes-facts-figures/

International Diabetes Federation. (n.d.). Type 1 diabetes. https://idf.org/es/about-diabetes/type-1-diabetes/

International Diabetes Federation. (n.d.). Type 2 diabetes. https://idf.org/es/about-diabetes/type-2-diabetes/

Jain, P., Joshi, A. M., Mohanty, S. P., & Cenkeramaddi, L. R. (2024). Non-invasive glucose measurement technologies: Recent advancements and future challenges. *IEEE Access, 12*, 61907–61936. https://doi.org/10.1109/ACCESS.2024.3389819

Jain, P., Pancholi, S., & Joshi, A. M. (2019). An IoMT based non-invasive precise blood glucose measurement system. In *IEEE International Symposium on Smart Electronic Systems* (pp. 111–116). https://doi.org/10.1109/iSES47678.2019.00034

Kajisa, T., Kuroi, T., Hara, H., & Sakai, T. (2024). Correlation analysis of heart rate variations and glucose fluctuations during sleep. *Sleep Medicine, 113*, 180–187. https://doi.org/10.1016/j.sleep.2023.11.038

Li, J., Tobore, I., Liu, Y., Kandwal, A., Wang, L., & Nie, Z. (2021). Non-invasive monitoring of three glucose ranges based on ECG. *IEEE Journal of Biomedical and Health Informatics, 25*(9), 3340–3350. https://doi.org/10.1109/JBHI.2021.3072628

Mardia, K. V., Kent, J. T., & Taylor, C. C. (1979). *Multivariate analysis*. Academic Press.

Naresh, M., Nagaraju, V. S., Kollem, S., Kumar, J., & Peddakrishna, S. (2024). Non-invasive glucose prediction using NIR technology with machine learning. *Heliyon, 10*(7), e28720. https://doi.org/10.1016/j.heliyon.2024.e28720

Okoye, K., & Hosseini, S. (2024). Analysis of variance (ANOVA) in R: One-way and two-way ANOVA. In *R programming: Statistical data analysis in research* (pp. 187–209). Springer. https://doi.org/10.1007/978-981-97-3385-9_9

Ramin, J., Mahyari, Z., Moulodi, M. J., Fatemi Ghiri, S. M., Tajalizadeh, H., Loloee Jahromi, A., Nakhostin, A., Abdollahifard, G., & Parsaei, H. (2021). An infrared non-invasive system for measuring blood glucose: A primary study using serum samples. *Biomedical Physics & Engineering Express, 7*(5). https://doi.org/10.1088/2057-1976/ac1c0b

Sun, X. (2022). Glucose detection through surface-enhanced Raman spectroscopy: A review. *Analytica Chimica Acta, 1206*, 339226. https://doi.org/10.1016/j.aca.2021.339226

Swapna, G., Soman, K. P., & Vinayakumar, R. (2018). Automated detection of diabetes using CNN. *Procedia Computer Science, 132*, 1253–1262. https://doi.org/10.1016/j.procs.2018.05.041

Swapna, G., Soman, K. P., & Vinayakumar, R. (2020). Diabetes detection using ECG signals: An overview. In *Studies in Big Data*. https://doi.org/10.1007/978-3-030-33966-1_14

Swapna, G., Vinayakumar, R., & Soman, K. P. (2018). Diabetes detection using deep learning algorithms. *ICT Express, 4*(4), 243–246. https://doi.org/10.1016/j.icte.2018.10.005

Townsend, K. L. (2024). One nervous system: Critical links between CNS and peripheral health. *Diabetes, 73*(12), 1967–1975. https://doi.org/10.2337/dbi24-0004

Zanelli, S., Ammi, M., Hallab, M., & El Yacoubi, M. A. (2022). Diabetes detection through PPG and ECG signals: A systematic review. *Sensors, 22*(13), 4890. https://doi.org/10.3390/s22134890

Zsombok, A., Desmoulins, L. D., & Derbenev, A. V. (2024). Sympathetic circuits regulating hepatic glucose metabolism. *Physiological Reviews, 104*(1), 85–101. https://doi.org/10.1152/physrev.00005.2023