© 2018 IJSRSET | Volume 4 | Issue 11 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology
DOI : https://doi.org/10.32628/IJSRSET2184111

Glucose Calibration Modeling in Blood with Spline Regression Approaching to Non-Invasive Tools

Ria Hayatun Nur*, Indahwati, Erfiani
Statistics Department, Bogor Agricultural University, Bogor, West Java,Indonesia

ABSTRACT

In this globalization era, health is the most important thing to be able to run various activities. Without good health, this will hinder many activities. Diabetes mellitus is one of the diseases caused by unhealty lifestyle.There are many treatments that can be done to prevent the occurrence of diabetes. The treatments are giving the insulin and also checking the glucose rate to the patients.Checking the glucose rate needs the tools which is safety to the body. This research want to develop non invasive tool which is safety and do not injure the patient. The purpose of this research is also finding the best model which derived from Linear, Quadratic, and Cubic Spline Regression. Some respondents were taking to get the glucose measuring by invasive and non invasive tools. It could be seen clearly that Spline Linear Regression was the best model than Quadratic and Cubic Spline Regression. It had 70% and 33.939 for R^2 and RMSEP respectively.

Keywords : Diabetes Mellitus, Glucose, Spline, Regression

I. INTRODUCTION

In this globalization era, health is the most important thing to be able to run various activities. Without good health, this will hinder various activities that will be carried out. Today, unhealthy lifestyles, such as consuming foods that contain high glucose levels, smoking, and rarely exercising are the causes of various diseases. One of the diseases that can be caused is diabetes mellitus.

Diabetes mellitus is one of serious diseases. This disease occurs when the pancreas is unable to produce enough insulin for the body or the body is unable to utilize insulin production properly [8]. There are two types of Diabetes Mellitus, namely type I Diabetes and Dibetes type II. Type I diabetes is diabetes caused by genetic factors. The damage is damage to pancreatic beta cells caused by genetic factors, immune power, and the environment. Nearly 5-10% of patients suffer from type I diabetes. The remaining 90-95% of people with diabetes are type II diabetes. In type II diabetes, there is a decreasing in insulin resistance or a decreasing in the amount of insulin production[1].The diabetes mellitus patients in 2015 reached 414.7 million and it was also predicted to increase to 641.7 million by 2040 [4].

There are several ways that can prevent to treat the patients with diabetes mellitus, one of which is giving the insulin to patients. Diabetes mellitus can come up at any time when the patient makes a mistake in his life style particularly on eating the food. The Control also can be done to prevent unwanted things from happening. One of the things is by checking the blood glucose. Usually, an invasive tool is using to check the blood glucose and it also can injure the body.

In the practice, using of invasive tools usually uses syringe. This is worried to endanger diabetics, which

was the difficulty of healing wounds which suffered by patients. Furthermore, other methods are needed to solve it. Therefore, this study tries to develop a tool for measuring blood glucose in the body by non-invasive blood glucose measuring devices without endangering them. When glucose in the blood is measured using a non-invasive tool, the tool will produce a spectrum of voltage values from the infrared.

This certainly requires a method that is able to model the output which produced by the spectrum in the form of residual intensity. It also needs the value as close as possible to an invasive blood measuring tool. In its application of modeling, there are two approaches used in modeling, namely the parametric method and the nonparametric method. The parametric method has limitations in estimating irregular changes in data patterns. If the assumptions which used in parametric modeling are unfulfilled, then the other alternative that can be used is the nonparametric method. The nonparametric method is the method that is subsequently selected in solving the problem of modeling glucose in the blood because of the form of data from intensity residues that do not form a particular pattern. One nonparametric method approach that can be used to model the output generated from infrared is calibration modeling using Spline Regression.

Spline regression is a method that uses a polynomial concept that has segmented properties. The concept is to fit the regression curves based on existing data so that the model will provide higher flexibility. The best model will be seen from the value of General Cross Validation(GCV) and the smallest Root Mean Square Error Prediction(RMSEP) [9].

In addition, problems regarding calibration have been studied in previous studies, namely the Bayes Approach [2], Wavelet [6], and Continuous Regression [5]. In addition, the Smallest Partial Squares Regression, Main Component Regression, and Artificial Neural Networks are also the methods used to overcome this problem [7].

II. METHODS AND MATERIAL

A. Data Source

The data used in this study are primary data from the Development and Prototype Clinical Test Data Monitoring Tool for Non-invasive Blood Glucose Levels. This data was collected from 118 respondents who were students of Bogor Agricultural University. This research conducted in December 2017 to April 2018 carried out by the biomarking non-invasive team which was spread from Department of Physics, Department of Statistics, Department of Nutrition Science, and Department of Computer Science. This research is one of the projects funded by the National Innovation System Research Incentive Program, Ministry of Research, Technology and Higher Education.

The independent variable used was the residual intensity value of the sensor output monitoring tool for non-invasive blood glucose levels. On the other hands, the dependent variable used was data on blood glucose levels with invasive measurements carried out by Prodia's clinical laboratory team having their address at Jl. Jend. Sudirman No. 38 B, Bogor 16143.

This data was obtained by measuring blood glucose levels using a non-invasive tool. Before the measurement, the respondents were measured their weight and their height. Then the sampling was carried out. In sampling, five repetitions were carried out assuming that in 15 minutes glucose levels in the response did not change. At the end of the sampling, respondents will measure the glucose in the blood using invasive methods, which was taking 4 ml blood sample on the veins in their arm. Furthermore, blood will be analyzed at the Prodia Clinic with a chemical process to determine blood glucose levels.

B. Methods

The steps were taken in this study as follows:

1. Exploring glucose data in the blood by looking the spreading of glucose levels in the blood with an invasive device.
2. Looking to the data structure on non-invasive blood glucose level monitoring tool, namely
   a. The pattern is formed from the time domain to residual intensity in 10 periods. This period known as modulation.
   b. After being identified, the last five modulations will be taken. Then the data will be determined at each peak and others will be cleaned.
   c. Furthermore, because the amount of data held by each peak is different, there will be a uniform amount of data with the selection of only three data from each shoot formed. The three data are initial data, middle data, and final data.
3. Then the data were prepared based on modulation, replication, and residual intensity.
4. The data which has been compiled, then will be selected for the best modulation of the five modulations
   a. The best modulation will be formed from the interaction plot between modulation, replication and residual intensity of each respondent.
5. After the best modulation is obtained, then the residual intensity value from the modulation of each respondent will be analyzed
6. Data for independent variables would be obtained from the average value of residual intensity in each respondent and the data of dependent variables were the results of measurements with invasive methods.
7. Divide data randomly into two groups, which was 80% for training data and the another was the testing data.
8. Then the data analyzed using Spline Regression.
   a. Model training data by previously determining the number of knots to be used.
   b. Building a Spline Regression model with various levels of orders, namely 1st order (linear), 2nd order (quadratic), and 3rd order (cubic)[3]

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_p x_i^p + \sum_{l=1}^{r} \beta_{l+m} (x_i - K_l)^m + \varepsilon_i \quad (1)$$

with y_i is the response to the first observation, β_0 is intercept, β_{l+m} is the slope in the explanatory truncated for 1st knot variable on the spline with the order m, K_l is the 1st knot, and r is the number of knots

c. Calculates the estimated values, Generalized Cross Validation (GCV) and R squared

$$GCV(k) = MSE(n^{-1} tr[I - A(k)])^2 \quad (2)$$

with Mean Square Error (MSE) (k) = $n^{-1} \sum_{i=1}^{n} (y_i - \hat{y})^2$, n is the number of observations, I is the identity matrix, and k is the point of knots $(k_1, k_2, \ldots, k_n)$, and A (k) = x (x'x) -1x'.

$$R^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 / \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad (3)$$

d. Data testing will be used as data for model validation. Then the data will be calculated with the estimated value, remaining, GCV, and Root Mean Square Error of Prediction(RMSEP).

e. Compare the minimum GCV results from the three models to get the best model

III.RESULTS AND DISCUSSION

A. Overview of Glucose Level With Invasive Measurement

There are 118 respondents. Figure 1 explains that the lowest blood glucose level of the respondent is 67 mg / dL while the highest blood glucose level is 276 mg / dL. On the other hands, the majority of respondents has blood glucose levels ranging from 75 mg / dL to 85 mg / dL and one of the respondents who has a highest glucose level . It is 276 mg / dL.

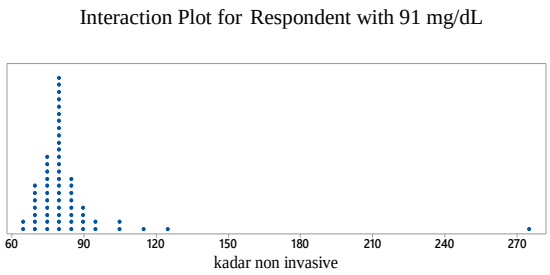

Figure 1: Blood glucose level with invasive measuring tool

B. General Description of Data with Non Invasive Measurements

Each respondents, five repetitions were carried out with the number of outputs produced by five periods. This period known as modulation. Modulation describes the level of lightning of a lamp. The modulation used in this study is modulation 50%, 60%, 70%, 80%, and 90%, meaning that there are five levels of lightning used in the study, namely 50%, 60%, 70%, 80%, and 90 %. If the modulation used is 60% modulation, then the lamp lives with a lightning level of 60%. Furthermore, the plot interaction between modulation and residual intensity based on replication in respondents with glucose levels of 91 mg / dL can be seen in Figure 2.

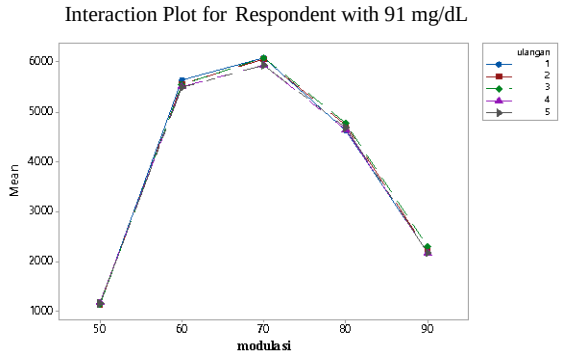

Figure 2 : Interaction plot for respondents with glucose 91 mg/ dL

Based on Figure 2,it showed that the research uses five modulations, 50% modulation is a modulation that has the lowest average residual intensity. At 60% and 70% modulation, there is an increasing percentage in the average residual intensity but when modulation is 80% and 90%, the residual value will decrease slowly. It means, by increasing modulation, the average residual intensity will increase, but this does not apply when the modulation used is 80% and 90% modulation. In addition, from the five replications used in each modulation, it can be seen clearly that 50% modulation is a modulation which has a stable loop rate. Stable repetition in this study has the meaning that the loop has an average value of residual intensity between replications that are not much different. This will cause the modulation standard value to be smaller. This is shown in Table 1.

TABLE 1.

GENERAL DESCRIPTION OF DATA ON RESPONDENTS WITH BLOOD GLUCOSE LEVEL 91 MG / DL

Variable	Modulation	N	Mean	St Dev
Residu	50	30	1159.7	48.3
Intensity	60	30	5531.5	98.5
	70	30	6012.4	98.0
	80	30	4681.1	84.9
	90	30	2179.9	115.5

Table 1 illustrates that the average value of residual intensity in respondents with a glucose level of 91 mg/dL. Its resulting that at 50% modulation is the smallest compared to other modulations, which is 1159.7. In addition, the standard deviation generated by 50% modulation has the smallest value of 48.3 compared to other modulation standard deviations. So it has a smaller level of diversity. The smallest level of diversity will be considered in modulation selection. Next, 50% modulation will be used for analysis.

C. Modulation 50% Data Overview

The data used for spline regression analysis is the average value of 50% modulation on each respondent. The explanatory variable is the residual intensity and the response variable is the value of the measurement of glucose in the blood by an invasive tool. In Figure 3 it can be seen clearly that the data does not create a particular function so it will be very difficult to model, so nonparametric regression is chosen to solve the problem.

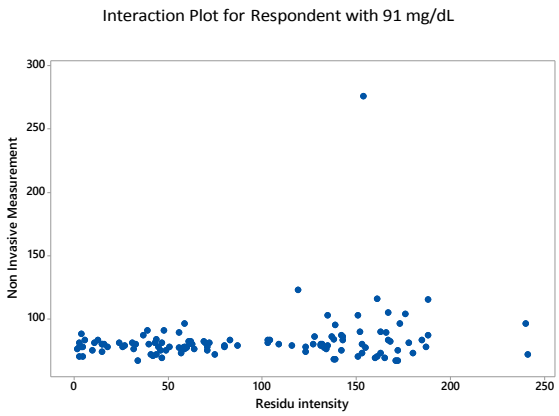

Figure 3: Data plot between the value of residual intensity and the measurement results of invasive tools

Nonparametric regression will be very helpful if the data are random and the shape of the curve is unknown. In an effort to make better modeling, the Spline Regression method is needed. Spline regression is able to accommodate all data by making it a curve that adjusts the data form. This form will create fragments in the curve.

D. Spline Regression

The spline regression used in this study will be applied to three different models, namely the Spline Linear, Quadratic Spline, and Cubic Spline models. The same treatment will be carried out on each model by simulating as much as 100 times the data that is available at each knot that is available. The software used is Software R 3.4.0.

Linear Spline would be marked with the first order, Quadratic Spline marked with second order, and Cubic Spline will be marked with third order. Then the best selection of knots will be seen from the minimum Generalized Cross Validation (GCV) value.

TABLE 2

RESULTS OF 100 TIMES REPETITION FOR SPLINE LINEAR, QUADRATIC AND CUBIC MODELS ON EACH KNOT

Ordo	Knot	GCV	Training RMSE	Testing RMSEP	R^2
1	8	106.585	9.526	33.939	0.702
1	9	107.276	9.49	36.042	0.707
1	10	106.988	9.409	38.544	0.716
2	8	267.502	14.882	79.658	0.441
2	9	267.412	14.777	68.267	0.456
2	10	267.944	14.708	87.007	0.467
3	8	222.877	13.562	755.298	0.514
3	9	215.587	13.479	939.395	0.526
3	10	210.678	13.097	253.759	0.553

Based on Table 2, GCV generated in first-order spline regression has the smallest GCV value compared to either Quadratic or Cubic Spline Regression. The smallest GCV value will be used in selecting the best knots. The best number of knots for full data is at knot 8. The largest R square value is 70 % also dominated by first order R square.

In addition, the graph bellow encourages the table above. It showed the relathionship between GCV and knot.

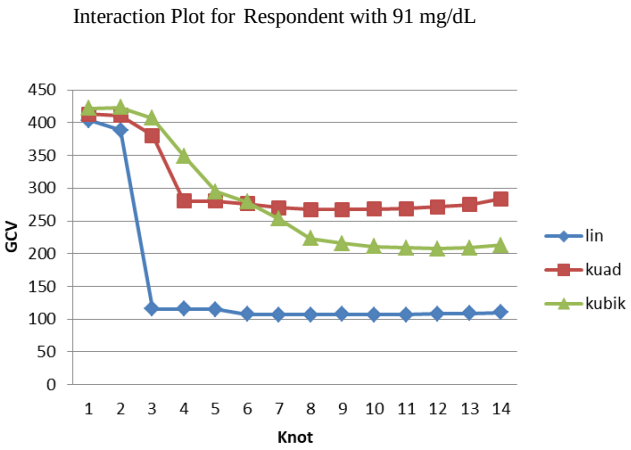

Figure 4 : Data plot between the value of GCV and knot

It can be seen clearly in Figure 4 that Spline Linear Regression is the best method compared to Spline Quadratic Regression and Spline Cubic Regression. It had the lowest GCV . It was almost 106.585 mg/dL.

IV.CONCLUSION

A. Appendices

Estimating the value of glucose in the blood using spline regression can determine the best model. It can be seen clearly from the estimation value which is close to the actual value. In addition, using linear Spline will be obtained a model with the smallest GCV, the largest R-square, and the smallest RMSEP. This component will make the basis that the linear model is the best model compared to the quadratic spline model, and the cubic spline model.

The model that can be used is

$$y = 90.2350 - 2.3543x + 2.3723(x - 4.990)_+ + 10.2530(x - 115.990)_+ - 13.8764(x - 119.009)_+ + 3.4653(x - 127.003)_+ + 77.4111(x - 151.995)_+ - 233.9294(x - 153.9996)_+ + 157.2210(x - 155.022)_+ - 0.9448(x - 187.890)_+$$

x : vector standard of deviation in intensity residu
y : vector glucose calibration measuring

B. Suggestion

The suggestion for the next researcher is the researcher can only take the best modulation for applying to the non invasive tools. Based on this research, the best modulation is 50 %. Besides, the sampling range of the glucose rate should be increased, it means the sample populations have the heterogenity in blood glucose.

V. REFERENCES

[1]. Brunner, Suddarth. 2010. Textbook of Medical-Surgical Nursing. Philadelphia: Wolters Kluwer Health.

[2]. Erfiani 2005 Pengembangan Model Kalibrasi dengan Pendekatan Bayes (Kasus Tanaman Obat) [disertation] (Bogor: Graduate Program, Institut Pertanian Bogor)

[3]. Eubank RL 1999 Nonparametric regression and spline smooting 2nd ed Marcel Dekker (New York)

[4]. IDF] International Diabetes Federation. 2015. IDF Diabetes Atlas Seventh Edition [Internet]. [download by 2018 March 15]. In: http://www.diabetesatlas.org/.

[5]. Setiawan 2007 Regresi Kontinum dengan prapemrosesan transformasi wavelet diskret. Jurnal ilmu dasar 8(2):103-109.

[6]. Sunaryo S 2005 Model Kalibrasi dengan Transformasi Wavelet sebagai Metode Pra-pemrosesan [disertation] (Bogor: Graduate Program, Institut Pertanian Bogor)

[7]. Tonah 2006 Pemodelan Kalibrasi Peubah Ganda dengan Pendekatan Regresi Sintyal P Spline [thesis] (Bogor: Graduate Program, Institut Pertanian Bogor)

[8]. WHO] World Health Organisation 2010 Diabetes [Internet] [download by 2018 March 26] In: www.who.int/nmh/publicatians/fact_sheet_diabe-tes_en.pdf

[9]. Wulandari et all 2017 Penerapan Analisis Spline untuk Menduga Harga Cabai di Jakarta Indonesian Journal of Statistics and Its Application Vol 1 No 1 (2017), 1 - 12.