Linear System Identification Models for Non-Invasive Glucose Level Estimation

Intan Maisarah Abd Rahim[1], Herlina Abdul Rahim[1*], Nur Athirah Syafiqah Noramli[2]

[1*] Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
[2] School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Shah Alam, Selangor. Malaysia

Corresponding author* email: herlina@utm.my
Available online 30 June 2024

ABSTRACT

Diabetes is a medical condition that can lead to severe health complications such as stroke, heart disease, blindness, and obesity. An estimated 347 million people worldwide were affected by diabetes, with approximately 3.4 million deaths attributed to high blood sugar levels. Researchers have explored various non-invasive techniques for measuring blood glucose levels, including ultrasonic sensors, multisensory systems, transmittance absorbance, bio-impedance, voltage intensity, and thermography. This paper examines the application of near-infrared (NIR) spectroscopy for glucose level measurement and the implementation of a linear system identification model to predict output data from NIR measurements. While NIR has been utilized in previous studies, there is ongoing debate regarding the optimal wavelength range, as different researchers have used varying wavelengths. To assess the feasibility of a linear approach, this study applies the Autoregressive Moving Average Exogenous (ARMAX) model to predict NIR measurement outcomes.

Keywords: Diabetes, Glucose level, Near-infrared (NIR), ARMAX.

1. Introduction

Glucose, commonly known as the primary carbohydrate, derives its name from the Greek word glykys, which translates to "sweet." It exists in two optical isomeric forms: D-glucose and L-glucose [1]. In the human body, glucose serves a critical function, and maintaining regular blood glucose monitoring is essential for diabetic patients to ensure levels remain within the normal clinical range of 3.5–6.1 mmol/L [2]. However, diabetes can arise because of inadequate dietary habits and lifestyle decisions. In 2004, around 347 million individuals globally were living with diabetes, and an estimated 3.4 million deaths were linked to complications caused by elevated blood sugar levels [3].

According to some references, blood glucose levels can be classified as follows [4]:
Diabetic classification:

$$Diabetic = \begin{cases} 0, & 4\ mmol/L \leq BGL \leq 7\ mmol/L \\ 1, & BGL > 7\ mmol/L \end{cases} \quad (1)$$

where:
0 = Non-diabetic individual
1 = Potential diabetic patient

For patients in critical condition, continuous diabetes management or monitoring is absolutely necessary. If glucose levels exceed the normal range, it is highly likely that the individual is suffering from diabetes. Continuous diabetes management can aid in maintaining blood sugar levels within the acceptable range for the human body, which is crucial to prevent the development of severe complications caused by elevated blood sugar levels. Over time, if diabetes is not properly controlled, it can lead to other serious health issues such as kidney failure, heart disease, and stroke [5].

In many instances, diabetic patients have expressed concerns about blood glucose testing, describing it as a painful process involving pinpricks and pinching. For critically ill patients, continuous monitoring often requires frequent finger pricking throughout the day, which can be both exhausting and painful. This process can be particularly challenging for patients diagnosed with dexterity limitations, aglophobia (fear of pain), or anxiety disorders. However, frequent

monitoring is essential for diabetic patients, especially those in severe or critical conditions, as diabetes is an incurable disease and a life-threatening metabolic disorder [6].

2. Methodology

This study involves data acquisition using both a conventional glucose meter as the reference device and near-infrared (NIR) spectroscopy as the input data source. The collected data was analyzed using a linear system identification model, specifically the Autoregressive Moving Average Exogenous (ARMAX) model, to evaluate its effectiveness in predicting blood glucose levels based on NIR wavelength data. Prior to analysis, the dataset went through a pre-processing stage to improve the quality and features of the input data.

2.1 Data Acquisition

The data was collected from three distinct groups: individuals diagnosed with diabetes, non-diabetic individuals, and a control group without any prior diagnosis. The data collection process was conducted under the supervision of personnel at the Outpatient Department (OPD) of Hospital Universiti Sains Malaysia (HUSM) in Kubang Kerian, Kelantan, Malaysia.

The human fingertips are considered one of the most suitable locations for measurement, alongside the earlobe, due to their thinner skin and higher concentration of observable blood vessels. The dataset comprises 135 sets of data obtained from the three subject groups. The data varies across factors such as gender, age, race, and medical condition (critical or non-critical). The study involving these subjects was carried out in compliance with the ethical principles outlined in the Good Clinical Practice guidelines and the Declaration of Helsinki [7,8]. Figure 1 illustrates the setup of the NIR spectrometer during the data measurement process.

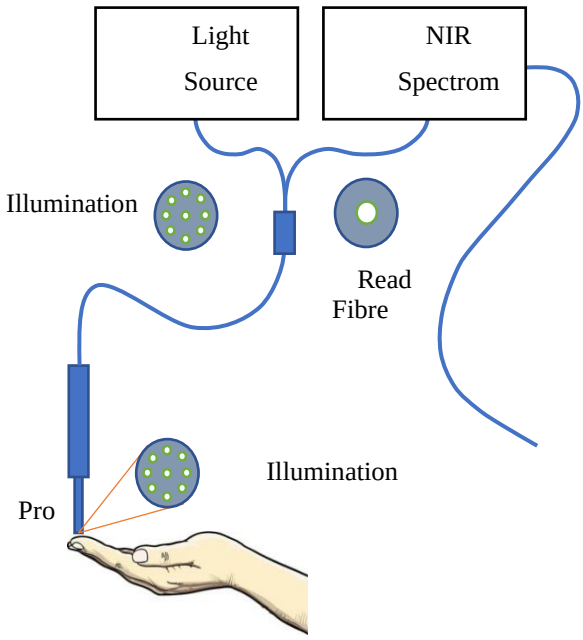


Figure 1. The illustration of the NIR spectroscopy instrument

2.2 Data Pre-processing

In this research, the data pre-processing methods employed include noise filtering, data sampling, interval correction, and data distribution. The Savitzky-Golay filter was utilized to refine the wavelength data by eliminating any unwanted noise or irrelevant information from the dataset. During the data sampling process, only the wavelengths containing useful information were extracted to serve as the input dataset. Subsequently, interval correction was applied to specific peaks within the NIR spectrum to further refine the data. Finally, the data distribution process was carried out to randomly divide the dataset into training, testing, and validation sets. Figure 2 illustrates the workflow of the data pre-processing stage.

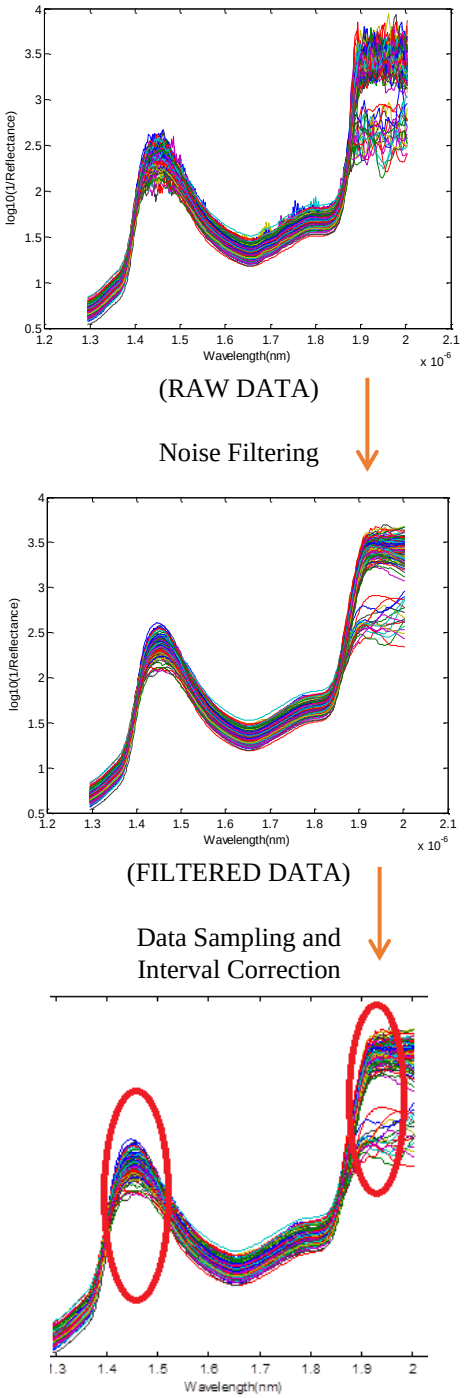


Figure 2. The flow of data pre-processing

3. Linear System Identification Model

In this study, the linear model used is the Autoregressive Moving Average Exogenous (ARMAX) model. Autoregressive Moving Average Exogenous (ARMAX) model is defined as:

$$G(z,p) = \frac{B(z)}{A(z)}, H(z,p) = \frac{1}{A(z)} \quad (2)$$

where $A(z) = 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a}, B(z) = b_1 z^{-1} + \cdots + b_{n_b} z^{-n_b}$, $G(p,z)$ and $H(p,z)$ are filters of finite order and functions of a parameter vector p,

$$p = [a_1 \ldots . a_{n_a} \ b_1 \ldots . b_{n_b}]^T \quad (3)$$

The un-regularized and regularized ARMAX models are both implemented to determine the optimum testing result of the system.

4. Model Validation

The Clarke Error Grid Analysis is a widely recognized method for evaluating the clinical accuracy of blood glucose level estimates compared to conventional invasive measurements [5,9–11]. This approach was utilized to validate the linear system identification model implemented in this research. The data from the output set and the reference set were plotted to identify the distribution regions [12] as shown in Table 1.

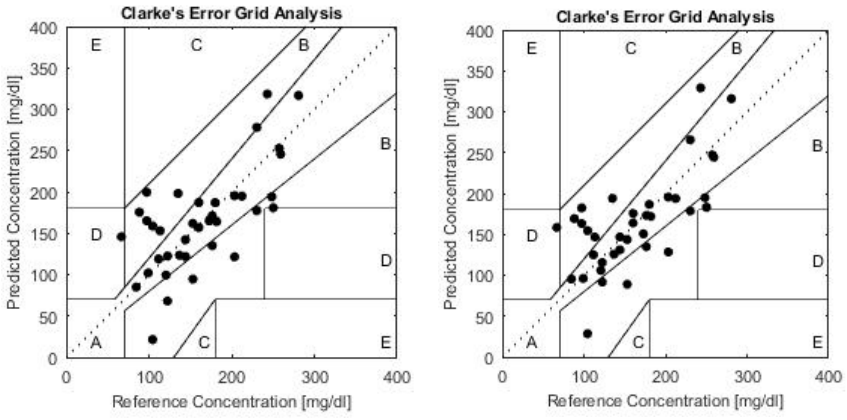
Table 1. Region distribution

Region	Description
Region A	within 20% of the reference sensor (clinically accurate).
Region B	outside of 20% but would not lead to an inappropriate treatment (clinically acceptable).
Region C	leading to unnecessary treatment (overcorrection).
Region D	indicating a potential dangerous failure to detect hypoglycaemia or hyperglycaemia (dangerous failure to detect).
Region E	confuse treatment of hypoglycaemia for hyperglycaemia and *vice versa* (serious error).

5. Results and Discussion

The prediction results of the ARMAX model, as shown in Table 2, indicate accuracy rates of only 47.79% for the unregularized model and 50.11% for the regularized model, both of which fall significantly short of the desired satisfactory outcomes.

During the modeling stage, a filter length of 33 was selected as the optimal value for training and testing. However, the results from the Clarke Error Grid Analysis were also unsatisfactory, as illustrated in Figure 3. The figure reveals that only 54.05% of the output data for the unregularized model (a) and 56.76% for the regularized model (b) fall within Region A, which represents the most accurate predictions.



(a) (b)

Figure 3. (a) The scattering of the un-regularized testing data (b) The scattering of the regularized testing data.

Table 2. The unregularized and regularized for training and testing result.

Filter Length	UNREGULARIZED						REGULARIZED					
	Training			Testing			Training			Testing		
	R	R^2	%	R	R^2	%	R	R^2	%	R	R^2	%
5	0.88	0.77	76.70	0.70	0.49	48.93	0.88	0.77	76.95	0.71	0.50	50.24
7	0.88	0.77	77.14	0.71	0.50	49.98	0.87	0.76	76.46	0.71	0.51	50.52
9	0.87	0.76	75.85	0.69	0.48	47.51	0.87	0.75	75.36	0.69	0.48	47.55
11	0.88	0.77	76.90	0.70	0.48	48.50	0.86	0.74	74.49	0.69	0.48	48.22
13	0.86	0.74	74.25	0.69	0.48	47.51	0.86	0.74	73.94	0.68	0.47	46.83
15	0.86	0.75	74.75	0.68	0.47	46.62	0.86	0.74	73.94	0.69	0.47	47.32
17	0.86	0.74	73.65	0.69	0.47	47.03	0.86	0.74	73.70	0.70	0.49	48.72
19	0.86	0.74	74.06	0.69	0.47	47.42	0.86	0.74	73.99	0.70	0.49	48.65
21	0.86	0.74	74.18	0.69	0.47	46.95	0.87	0.75	74.82	0.70	0.49	48.54
23	0.86	0.74	73.96	0.67	0.45	45.36	0.87	0.75	74.89	0.69	0.47	46.95
25	0.86	0.74	74.20	0.67	0.45	45.33	0.87	0.75	75.10	0.69	0.47	47.18
27	0.87	0.76	75.78	0.68	0.47	46.64	0.87	0.76	76.02	0.70	0.49	48.69
29	0.87	0.76	76.02	0.69	0.48	47.67	0.88	0.77	76.67	0.71	0.50	50.01
31	0.88	0.77	76.70	0.69	0.48	48.09	0.88	0.77	77.11	0.71	0.50	49.86
33	0.87	0.77	**76.55**	0.69	0.48	**47.79**	0.88	0.77	**76.65**	0.71	0.50	**50.11**
35	0.88	0.78	77.93	0.73	0.53	53.23	0.87	0.77	76.51	0.76	0.57	57.06
37	0.88	0.77	77.09	0.73	0.53	52.58	0.87	0.76	75.79	0.73	0.53	53.42

6. Conclusion

This paper discusses the implementation of the linear system identification model, which, however, demonstrates the ineffectiveness of the ARMAX model in predicting the output values of the NIR spectrum. The unsatisfactory results may be attributed to various environmental and physiological factors during NIR measurements, such as human body temperature, ambient temperature, humidity, as well as the skin color and texture of the patients. To enhance prediction accuracy, future work is proposed to explore the implementation of nonlinear models, aiming to expand the range of investigated parameters and improve the overall results.

Acknowledgment

The authors would like to express our deepest gratitude to all parties who have contributed to this research, especially Universiti Teknologi Malaysia and the Outpatient Department, Hospital Universiti Sains Malaysia (USM), Kubang Kerian Kelantan for their utmost cooperation.

References

[1] T. Denisova, L. Malinova, *Glucose, in: Handb. Opt. Sens. Glucose Biol. Fluids Tissues*, Taylor & Francis, pp. 1–40, 2008, doi:doi:10.1201/9781584889755.ch1.

[2] E.M.A. A. Trabelsi, M. Boukadoum, C. Fayomi, "Blood glucose sensor implant using NIR spectroscopy: Preliminary design study", in: Microelectron. (ICM), *2010 Int. Conf. On, Cairo*, 2010: pp. 176–179, 2010.

[3] C.I. Examinations, "Facts and figures about diabetes", (2016) 2016. http://www.cie.org.uk/images/268776-facts-and-figures-about-cambridge-international-examinations.pdf (accessed January 01, 2014).

[4] C.F. So, K.S. Choi, J.W.Y. Chung, T.K.S. Wong, "An extension to the discriminant analysis of near-infrared spectra", *Med. Eng. Phys.* 35,172–177, 2013, doi:10.1016/j.medengphy.2012.04.012.

[5] L. Ben Mohammadi, T. Klotzbuecher, S. Sigloch, K. Welzel, M. Göddel, T.R. Pieber, et al., "In vivo evaluation of a chip based near infrared sensor for continuous glucose monitoring", *Biosens. Bioelectron*. 53 (2014) 99–104, 2014, doi:10.1016/j.bios.2013.09.043.

[6] S. Sivanandam, M. Anburajan, B. Venkatraman, M. Menaka, D. Sharath, "Estimation of blood glucose by non-invasive infrared thermography for diagnosis of type 2 diabetes: An alternative for blood sample extraction", *Mol. Cell. Endocrinol*. 367, 57–63,, 2013, doi:10.1016/j.mce.2012.12.017.

[7] C. Review, S. Communication, G. "Principles, World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects", *Nurs. Ethics*. 9, 105–109, 2002.

[8] A. Caduff, M. Mueller, A. Megej, F. Dewarrat, R.E. Suri, J. Klisic, et al., "Characteristics of a multisensor system for non invasive glucose monitoring with external validation and prospective evaluation", *Biosens. Bioelectron*. 26, 3794–3800, 2011, doi:10.1016/j.bios.2011.02.034.

[9] A. Caduff, M.S. Talary, M. Mueller, F. Dewarrat, J. Klisic, M. Donath, et al., "Non-invasive glucose monitoring in patients with Type 1 diabetes: A Multisensor system combining sensors for dielectric and optical characterisation of skin", *Biosens. Bioelectron*. 24, 2778–2784, 2009. doi:10.1016/j.bios.2009.02.001.

[10] E. Monte-Moreno, "Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques", *Artif. Intell. Med*. 53, 127–138, 2011. doi:10.1016/j.artmed.2011.05.001.

[11] M. Aloraefy, J.T. Pfefer, C.J. Ramella-Roman, E.K. Sapsford, "In Vitro Evaluation of Fluorescence Glucose Biosensor Response", *Sensors* . 14, 2014. doi:10.3390/s140712127.

[12] W.L. Clarke, D. Cox, L. a Gonder-Frederick, W. Carter, S.L. Pohl, "Evaluating clnical accuracy of systems for self-monitoring of blood glucose", *Diabetes Care*. 10, 622–628, 1987.