International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 02, 2020
ISSN: 1475-7192

TIME SERIES PREDICTION ARMA MODEL FOR PREDICTING BLOOD GLUCOSE IN ARTIFICIAL PANCREAS

[1]CIFHA CRECIL DIAS, [2]SUREKHA KAMATH, [3]SUDHA VIDYASAGAR

ABSTRACT--- *Patients with diabetes requires continuous monitoring of blood glucose level. Over the past few decades, Continuous glucose monitoring (CGM) has become a very helpful tool to manage and record glucose levels in the blood. With the help of CGM, the control and regulation of blood glucose can be achieved. Collecting the data from CGM, the paper attempts to predict future glucose levels by applying Auto Regressive Moving Average (ARMA) model. This predicted glucose level can be used for forecasting, and immediate appropriate action can be employed to avoid the risks related to diabetes. A good efficient model and a controller can be developed to improve the control using the Model Predictive Control technique with the predicted data set. The major risks that can be avoided are hyperglycemia and hypoglycemia. This paper elaborates on the estimation and prediction of blood glucose levels using the ARMA model in MATLAB platform. The CGM data is collected from a type 1 diabetic patient, and five-day data is recorded using the CGM device. The time series of the collected raw data is analyzed, and the parameter estimation is obtained. The model order is selected, and forecasting models are determined. This type of method for prediction gives good prediction with a lesser error when compared with original raw data and estimated values.*

KEYWORDS-- *ARMA model, Diabetes, Prediction, Model estimation and Control*

I. INTRODUCTION

Being the second most populated country in the world, India is facing major health-related issues, among which Diabetes is rapidly growing. As the lifestyle of the people is changing with the technology that seems to make the quality of life the man lives is also changing. With the increase in the complications related to diabetes and the cost of affording medication is lakhs together to get better control and monitoring while keeping aside the regular insulin injections. It is important to maintain blood glucose in a very narrow range, say 70-110 mg/dl [1]. Many external disturbances affect the levels of blood glucose and lead to various complications. The main control of the blood glucose is through the pancreas which releases hormones like insulin and glucagon, which regulates the level of blood glucose. Among types of diabetes Type 1 is autoimmune, which destroys the Beta-cell, and the production of insulin is nullified. Such patients depend on the external injection of insulin to maintain blood glucose [2].

[1] *Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal Karnataka India 576104, cifhasaldanha@yahoo.in*

[2] *Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal Karnataka India 576104, surekha.kamath@maipal.edu*

[3] *Department of Medicine, Kasturba Medical College, Manipal Academy of Higher Education, Manipal Karnataka India 576104, vsagar33@yahoo.com*

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

In several clinics and hospitals, the glucose management and diagnosis of blood glucose level in the body is still determined using the finger-prick technique. This should be done multiple times a day for a Type 1 patient especially in ICU (Intensive Care Unit), to keep track and avoid complications [3],[4]. Though CGM devices are available, the patient is not educated enough for self-monitoring; to avoid complications and risks, prediction models can be developed to predict the future glucose level which helps the diabetes patient to control and take action. Such prediction models moderately give a future prediction, and this is done with the help of recorded CGM data [5],[6].

The glucose concentration in the human body changes over time and this can be considered as data series and can be taken for prediction as time-series data. Hence considering a time series method for the analysis of the prediction will give a reasonable approach. Several types of research have been carried out considering the time series data especially for developing artificial pancreas with new technologies like Artificial Neural Network (ANN), Machine Learning, multi/ mixed technique based on CGM data. ANN methods are used for the identification of the dynamics of the patient based on the insulin dosage. Machine learning also requires a variety of inputs and insulin concentration, and food intake to predict. Moreover, for multi and mixed techniques, the combination of linear and nonlinear mainly used to reduce the complication of hyper and hypoglycaemia [7]. The range of hyper and hypoglycaemia is described below, where the normal range is considered to be 70-110 mg/dl [8],[9]

- Hyperglycaemia is a situation where an excessive amount of blood glucose starts to circulate in the blood and is identified in two categories:
- Fasting hyperglycaemia is a condition where the blood glucose rises above 130 mg/dl for 8 hours of fasting.
- Postprandial hyperglycaemia is a condition where the blood glucose rises above 180mg/dl after 2 hours of meal intake.
- Hypoglycaemia is a state where blood glucose becomes lower than the acceptable range and is categorized into two terms:
- Slight hypoglycaemia where the blood glucose range is 55-70 mg/dl$.
- Severe hypoglycaemia where the blood glucose range is below 55mg/dl.

The ANN, Machine learning models, requires multiple knowledge like dosage of insulin, food intake, the concentration of insulin, etc for an effective prediction. The data collection for such models entirely depends on the single subject inputs provided and the subject may not be a professional user to read the data collected [7]. For a broader application, such data collected might be challenging to interpret too. Considering these limitations, we have aimed a practical model to predict the blood glucose, by just examining the recorded or historical saved CGM data.

The blood glucose levels are measured every 5 minutes and are recorded using the CGM device as a fixed instant of time. This data is considered as the historical data, and prediction is implemented by estimating the model order and the parameter for forecasting models. The method used in research is the ARMA model; such a technique requires only output information that is the previously saved glucose measurement. The model orders are determined by the Akaike Information Criterion (AIC), which helps in forecasting hypo and hyperglycaemia. The Recursive Least Square (RLS) is used for the parameter estimation, and then the prediction models are developed for forecasting [10].

This paper elaborates on the ARMA model application which is used for the prediction of blood glucose in a diabetic subject. The paper is structured into sections; Section 2 contains Methods and Methodology, Section 3 elaborates the results obtained, followed by concluding remarks in Section 4.

II. METHODS AND METHODOLOGY

In this section Time Series Prediction Method is used for prediction and forecasting is briefly presented using the ARMA model for the blood glucose prediction.

A. Analysis of Time Series Prediction Method:

Empirical models do not take into account any information about real-time or accurate models. These models can be developed with only the knowledge of collected raw output data, and the parameters can be estimated from the data collected. The simplest and the suitable model available for such data is ARMA (Autoregressive Moving average) model [11]."An ARMA model is used to describe the stochastic time series in terms of two polynomials. The first of these polynomials is for autoregression, the second for the moving average. This model is referred to as the ARMA (p, q) model; where is the order of the autoregressive polynomial, q is the order of the moving average polynomial. AR(p) makes predictions using previous values of the dependent variable. MA(q) makes predictions using the series mean and prior errors". The typical flow diagram process for a time series prediction is shown in Figure 1 and the steps are explained below,

Step 1: The raw data is collected from CGM, which is recorded for every 5 minutes. the cross validation is done by leave-one-out procedure dividing the data into training and validation sets.

Step 2: Stationarity test: The test was carried with Unit root test, to remove the non-stationary. The backward difference operator was considered only if such a situation was present. Sample autocorrelation function (ACF) and partial (PACF) was conducted.

Step 3: Model order: ACF and PACF were conducted to check the order of the model, especially of Autoregressive AR and Moving average MA terms.

Step 4: Parameter **Estimation:** The estimation of the parameter is done by maximum likelihood, and AIC is used for model selection.

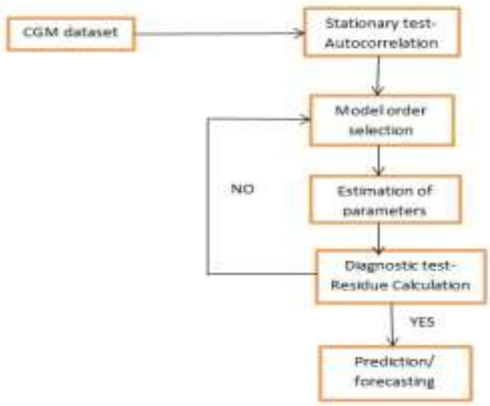

Fig. 1 Steps to build a Prediction Model

Step 5: Diagnostic test: The plots of ACF and PACF are plotted to analyze the confidence level through autocorrelation.

Step 6: Forecasting: The predicted model is forecasted by checking the Mean Absolute Error.

B. Prediction ***Based on ARMA Model:***

The most frequently used model for describing a stationary time series is the ARMA model. The detailed explanation for building a prediction model for glucose concentration is highlighted in this section.

Step 1: Model Description

In the proposed work we have considered a diabetic patient's blood glucose reading as a random time series. With the ARMA models, the future glucose readings or the current glucose reading is given by the function of previous glucose reading as in equation 1 [12],[13]:

$$Y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t \qquad (1)$$

Where y is the measurement of glucose reading at current time t, y_{t-i} is the reading i times before the current time or previous readings, $a_1, a_2 \ldots\ldots a_p$ are the Auto Regressive AR parameters, p is the AR model order, $b_1, b_2 \ldots\ldots b_q$ are the Moving average MA parameters with q as the MA model order. In equation 1 a_i, b_i are not known, they can be identified from the glucose readings. ε_t is the white noise which contains identical distribution with y_t and is independent, where

$$E(\varepsilon_t) = 0,\ var(\varepsilon_t) = \sigma^2 > 0 \qquad (2)$$

Step 2: Steps for Prediction

Let $y_1, y_2 \dots . y_t$ be the time series glucose reading collected from a diabetic patient. The entire collected data is plotted over time to check the periodicity of time series. The time series will be divided with period length if the series is periodic. The zero mean normalize for the time series is obtained by equation 3, where $\overline{y}$ is the mean of y_t:

$$\boldsymbol{\grave{y} = y_t - \overline{y}} \qquad (\mathbf{3})$$

The ACF and PACF are evaluated and the confidence interval is considered as $\frac{-2}{\sqrt{n}}, \frac{2}{\sqrt{n}}$, where n is the number of data of $\grave{y}_t$. This is assessed to check if the $\grave{y}_t$ is a stationary data set or non-stationary set. The series is said to be stationary if autocorrelation lies within the confidence interval. The ARMA model is used for prediction only if the series is stationary.

Step **3:** Parameter **Estimation**

Depeding on the obtained autocorrelation the model parameters and the model order for AR(p) and MA(q) can be determined by:

- if AIC is decreased at q then p=0 the model is MA(q)
- if AIC is decreased at p then q=0 the model is AR(p)
- if AIC is not decreased at any point then the model is ARMA (p, q)

Using System Identification and by coding in MATLAB 2017 platform, the model order for ARMA (p, q) can be determined. To select (p, q) order for each AIC is used and is defined by equation 4:

$$AIC = \frac{1}{n}(-2L + 2k) \qquad (4)$$

Where L is the value of Likelihood, k is the number of parameters and n is the number of observations. The minimum combination of AIC $0 \geq p \geq \sqrt{n}$ and $0 \geq q \geq \sqrt{n}$ is considered. The prediction formula can be written as in equation 5 and the Mean absolute error is given in equation 6 [14],[15]:

$$\grave{y}_t = a_1 \grave{y}_{t-1} + \cdots + a_p \grave{y}_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q} \quad (5)$$

$$Mean\ Absolute\ Error = \frac{1}{n}\sum_{i=1}^{n} |e_i| \qquad (6)$$

The equation 5 gives the predicted future glucose values and can be compared with the original raw data to find the error. Higher the data size the accuracy of the prediction also increases.

III. RESULT AND CASE STUDY

The glucose reading of a diabetic patient is collected from an open data base source of an hospital, which is monitored using CGM with 5 minutes as time interval for 5 days. The graphical visualization of the raw data sample is shown in figure 2,3,4,5 and 6. These figures have x-axis as index which is time interval, every instance is 5 minutes, y-axis the blood glucose measurement measured in md/dl [16], similarly five day data is plotted.

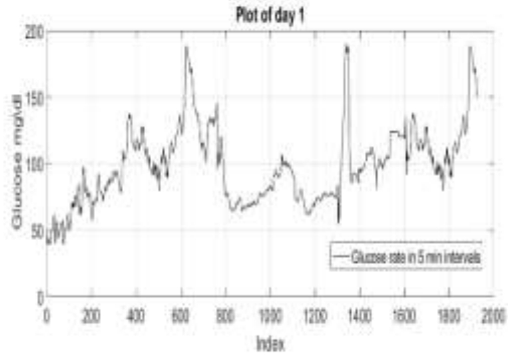

Fig. 2 Plot of day 1 data

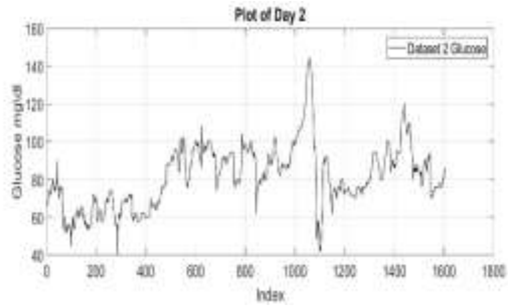

Fig.3 Plot of day 2 data

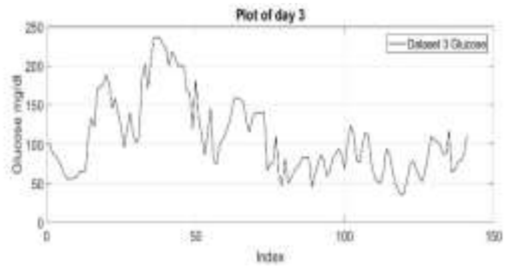

Fig. 4 Plot of day 3 data

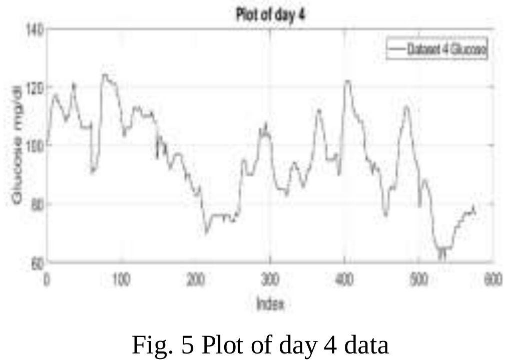

Fig. 5 Plot of day 4 data

International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 02, 2020
ISSN: 1475-7192

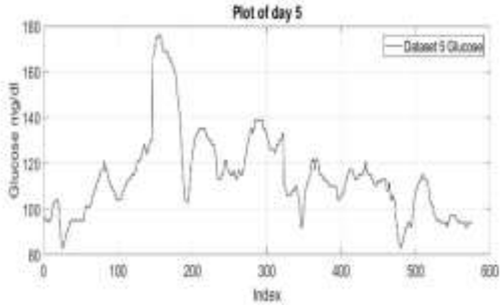

Fig. 6 Plot of day 5 data

The collected data is not periodic; hence zero mean normalization has been attained. The autocorrelation and partial autocorrelation is estimated using MATLAB and is shown in figure 7.

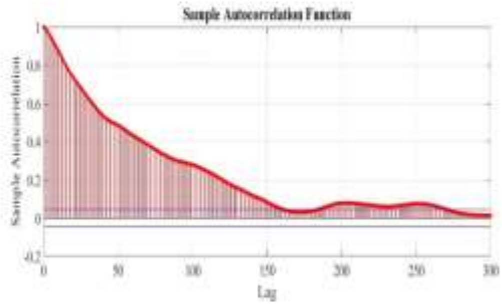

Fig.7 Auto Correlation plot

The autocorrelation in figure 7 has significant values up to lag 160 and this means the data at the current instant depends on the past data values at approximate 160 samples. Each sample is at any instant of five minutes, the autocorrelation function is within the confidence interval and can be predicted using the ARMA model. Using the best fit technique, the structure of the model is chosen to be ARMA (6,0,6), this is selected according to the estimation error of different models in MATLAB as shown in figure 11 . Firstly, consider p=q=1, calculate the AIC for ARMA (1,1), further check for the combination until the findings of AIC is minimum. The ARMA (6,6) has the least error and it is selected for further prediction. The best model fit is shown in table 1 with the model structure (6,6), the equation of the model can be written as in equation 7;

$$y_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi y_{t-i} + \sum_{j=1}^{q} \theta \varepsilon_{t-j} \qquad (7)$$

$$y_t = 0.834 + 6.705e^{-05} - 0.113y_t + 1.5y_{t-2} + 0.605y_{t-3} - 0.795y_{t-4} - 0.378y_{t-5} + 0.113y_{t-6} + 1.33\varepsilon_{t-1} + 0.068\varepsilon_{t-2} - 0.78\varepsilon_{t-3} - 0.44\varepsilon_{t-4} - 0.158\varepsilon_{t-5} - 0.0915\varepsilon_{t-6}$$

The best model fit is shown in table 1 with the model structure (6,6); the equation of the model is written as in equation 7 with the parameters. From this equation, $E(\varepsilon_t) = 0$ and the variance is 7.38006. y_{t-k} Is the zero-mean normalization of the measurement of blood glucose time series is $\varepsilon_t$created by MATLAB. Figure 8 shows the estimated model and the original data, and it can be deduced that the estimation accuracy is almost close to the original data, only a few points when the blood glucose changes suddenly the estimation error has been high. This is because the human body is a slowly varying system which response with some delay. The overall estimated error is given by equation MSE (Mean square error) and figure 9:

Model order	AICBIC	AR Coeff	MA Coeff
ARMA (6,6)	159.187	6	6

Table 1. Model order selection table

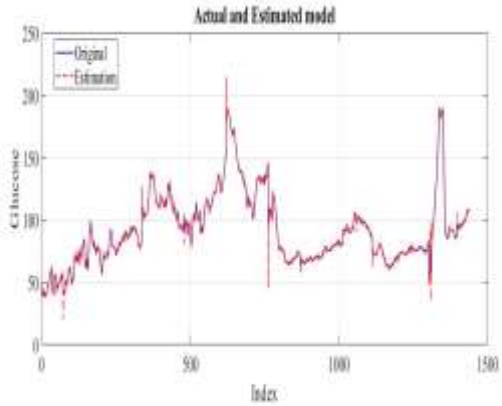

Fig 8. Estimation and original data ARMA model

$$MSE = \frac{1}{n}\sum_{i=1}^{n}|e_i| = 6.705e^{-05}$$

The prediction using the five data model for 5 minutes to 40 minutes is plotted, it is observed that up to 20 minutes after the day 5, the prediction accuracy is good, and the prediction error is low. Such models can be used to predict the blood glucose to avoid the hyper and hypoglycemia risks associated with the diabetic patient. The prediction or forecasting model is shown in figure 9, the original data set is plotted with the predicted dataset and the results can be compared by seeing the error obtained by the same. The prediction is made for 20 minutes, 30 minutes and 40 minutes, one step ahead prediction.

The diagnostic check is done by the autocorrelation plot for the residue to analyses the test for significant spikes. The sample at zero should be at one and the remaining samples should remain within the boundary. Figure 10 shows the autocorrelation of the residue samples.

International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 02, 2020
ISSN: 1475-7192

The figure 11 is directly taken from the MATLAB command window, which shows the difference between the ARMA model with order one and order 6, the order 6 has lesser variance compared to simple order 1, and the parameters of the model are estimated for the model development.

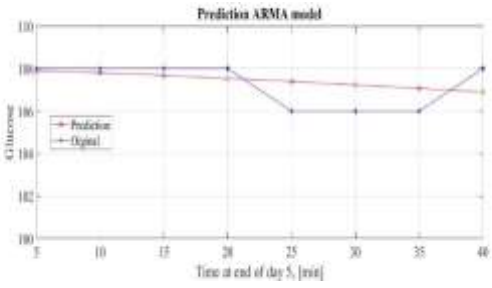

Fig 9. Forecasting prediction comparison with original and estimated

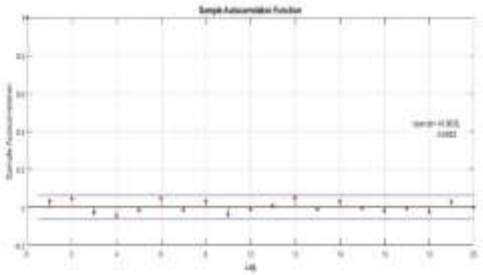

Fig 10. Autocorrelation of Residue sample

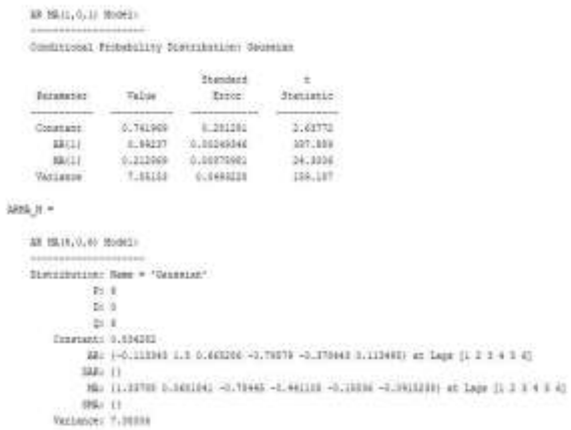

Fig 11. Parameter Estimation ARMA model

IV. CONCLUSION

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

ARMA model is basically used for the time series analysis method to quote a few advantages: it requires less prior information about the time series; its calculation process is accurate [17],[18]. Hence such a technique is used to develop a prediction model in our research. In this research, the application of this model is used for blood glucose prediction for diabetic patients, which gives an excellent performance and satisfying result for one step ahead prediction. Several types of research have been conducted in developing an adequate diabetic patient model using blood glucose readings from CGM. Though the reading is collected through CGM, a few disadvantages, like many patients may not record their blood glucose continuously due to this, the model developed may be limited to available data set reducing the accuracy.

For the developed model, an alarm can be set up for further future work to avoid a sudden increase in the decrease in blood glucose reading. With the addition to this, a controller can be developed, especially using the Model Predictive technique to regulate and control the blood glucose. Meanwhile, the results of the presented work bring it very close to the development of a device that can measure and control the blood glucose accurately.

REFERENCES

1. S. Coster, M. Gulliford, P. Seed, J. Powrie, R. Swaminathan, "Monitoring blood glucose control in diabetes mellitus: a systematic review", *British Journal of Clinical Governance-Bradford* 5 (4) (2000) 225–227.
2. R. Pradeepa, V. Mohan, "Prevalence of type 2 diabetes and its complications in India and economic costs to the nation" *European Journal Of Clinical Nutrition* 71 (7) (2017) 816.
3. P. Zhang, E. Gregg, "Global economic burden of diabetes and its implications" *The Lancet Diabetes & Endocrinology* 5 (6) (2017) 404–405.
4. W. H. Herman," The global burden of diabetes: an overview, in: Diabetes mellitus in developing countries and underserved communities" *Springer*, 2017, pp. 1–5.
5. D. Rodbard," Continuous glucose monitoring: a review of recent studies demon- starting improved glycaemic outcomes" *Diabetes Technology & Therapeutics* 19 (S3) (2017) S–25,
6. D. C. Klonoff, D. Ahn, A. Drincic, "Continuous glucose monitoring: a review of the technology and clinical use*", Diabetes Research And Clinical Practice* 133 (2017) 178–192.
7. J. Yang, L. Li, Y. Shi, X. Xie, "An arima model with adaptive orders for predicting blood glucose concentrations and hypoglycemia,"*'IEEE Journal Of Biomedical And Health Informatics* 23 (3) (2018) 1251–1260.
8. J. Wang, J. Zgibor, J. T. Matthews, D. Charron-Prochownik, S. M. Sereika,L. Siminerio, "Self-monitoring of blood glucose is associated with problem- solving skills in hyperglycemia and hypoglycemia", *The Diabetes Educator* 38 (2) (2012) 207–218.
9. E. Daskalaki, K. Nørgaard, T. Zu¨ger, A. Prountzou, P. Diem, S. Mougiakakou, "An early warning system for hypoglycaemic/hyperglycaemic events based on fusion of adaptive prediction models", *Journal Of Diabetes Science And Technology* 7 (3) (2013) 689–698.
10. Z. Ying, K. Rui, X. Shihong, "Research on glucose concentration predicting based on arma model", *2014 Prognostics and System Health Management Conference (PHM-2014 Hunan), IEEE*, 2014, pp. 332–335.

11. T. Singye, S. Unhapipat, “Time series analysis of diabetes patients: A case study of jigme dorji wangchuk national referral hospital in Bhutan”, *Journal of Physics: Conference Series,* Vol. 1039, IOP Publishing, 2018, p. 012033.
12. E. Montaser, J.-L. D´ıez, J. Bondia, “Stochastic seasonal models for glucose prediction in the artificial pancreas”, *Journal Of Diabetes Science And Technology* 11 (6) (2017) 1124–1131.
13. I. Contreras, S. Oviedo, M. Vettoretti, R. Visentin, J. Veh´ı, “Personalized blood glucose prediction: A hybrid approach using grammatical evolution and physiological models”,*PLOS ONE* 12 (11) (2017) e0187754.
14. S. Fiorini, C. Martini, D. Malpassi, R. Cordera, D. Maggi, A. Verri, A. Barla, “Data-driven strategies for robust forecast of continuous glucose monitoring time-series”, 2017 39th *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE*, 2017, pp. 1680–1683.
15. M. Eren-Oruklu, A. Cinar, L. Quinn, D. Smith, “Estimation of future glucose con- centrations with subject-specific recursive linear models”, *Diabetes Technology & Therapeutics* 11 (4) (2009) 243–253.
16. G. Sparacino, F. Zanderigo, S. Corazza, A. Maran, A. Facchinetti, C. Cobelli, “Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series”, *IEEE Transactions on biomedical engineering* 54 (5) (2007) 931–937.
17. J. Reifman, S. Rajaraman, A. Gribok, W. K. Ward, “Predictive monitoring for im- proved management of glucose levels”, *Journal Of Diabetes Science And Technology* 1 (4) (2007) 478–486.
18. F. Sta˚hl, R. Johansson, “Diabetes mellitus modeling and short-term prediction based on blood glucose measurements”, *Mathematical Biosciences* 217 (2) (2009) 101–117.