Application of Machine Learning Methods in the Blood Glucose Prediction
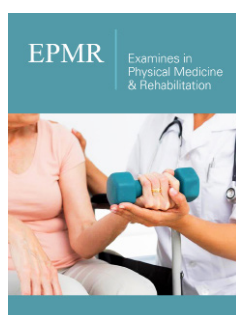
Artur Wodołażski*

Department of Energy Saving and Air Protection, Poland

Abstract

Type 1 diabetes (T1D) is a chronic disease that requires patients to know the blood glucose values to ensure their normal levels. The methods for predicting blood glucose are one of the areas of interest for clinical researchers. The literature describes many methods of predicting blood glucose, which require the determination of many activities, such as the time of insulin injection or emotional factors that may be susceptible to errors. To reduce the impact of individual activities, continuous glucose monitoring (CGM) is proposed to predict the level of glucose in the blood regardless of other factors. The paper presents the comparison of machine learning methods such as: linear regression, vector (SVR) and selected evolutionary algorithms (AE) to predict glucose concentration. The proposed methods based on SVR, and AE algorithms have achieved high accuracy and quality of predicted results.

Keywords: Machine learning; Evolutionary algorithms; Glucose concentration

Introduction

In the human body, the regulation of blood glucose levels is controlled by the action of two hormones: glucagon and insulin. The loss or destruction of β cells in the pancreas causes type 1 diabetes (T1D). Therefore, a reduction in insulin production leads to an increase in blood glucose and hyperglycemia [1]. Insulin plays a significant role in the ability of cells to metabolize glucose [2]. Hyperglycemia is usually associated with complications such as long-term microvascular complications (diabetes, neuropathy, retinopathy, and nephropathy) that cause cardiovascular problems (stroke, peripheral arterial disease, and coronary artery disease). In addition, a decrease in blood glucose levels hypoglycemia can quickly turn into a critical situation for the patient, such as nervousness, sweating, rapid heartbeat, headaches and even coma. In line with the latest technological advances in diabetes.

Patients have the ability to monitor their blood glucose levels continuously (every few minutes). The devices used (CGM - Continuous Glucose Monitoring) represent a continuous method of monitoring blood glucose levels [3]. These sensors are placed subcutaneously in the patient's interstitial fluid. In fact, the lag time between blood glucose and interstitial fluid can be as long as 5 minutes. This time is needed for the diffusion of the sensor and the computation of the time delay due to the filtering operation used to measure the CGM. In order to shorten the response and response times of the sensor, new efficient algorithms are needed that are able to predict the concentration of glucose over a given time horizon.

Machine learning techniques are often used in biology, image analysis or path morphology [4]. They are divided into supervised and unsupervised learning. Machine learning algorithms with the teacher include the support vector method, regression methods and neural networks. For tutorless learning methods, for example, k-means algorithms, auto-coders. Mathematical modeling includes methods that can be applied and implemented in electronic devices. However, in practice, they did not meet performance expectations due to limited precision and

dependence on patient activity measurements [5]. Bremer et al. [6] tested blood glucose levels, recording data every 10 minutes. The experimental results obtained from the modeling of blood glucose data showed a time horizon of 10 minutes. Studies have shown that blood glucose levels can be predicted by using past blood glucose values. Sparacino et al. [7] compared the predictive accuracy of the autoregressive model (ARM) with the first-order polynomial model. The results showed that the ARM model was the most consistent for obtaining significant performance at 30-45 minutes ph. Palerm et al. [8] used the Kalman filter to predict blood glucose levels based on a derivative reconstruction of glucose levels.

Pappada et al. [9,10] proposed an artificial neural network (ANN) written on the basis of the "NeuroSolutions" software package, which allowed to predict blood glucose levels for a time horizon of 50-180 minutes. The experimental results showed that the predicted blood glucose levels were more accurate in the hyperglycemic and normoglycemic stages than in the hypoglycemic stages. Consequently, it can be concluded that the ANN prediction results largely depend on the quality and set of training data for these networks. The article presents the possibilities of using selected machine learning techniques in predicting the patient's blood glucose. The goal is to accurately predict the blood glucose level of T1D patients using only CGM data as input to the learning phase. The practical implementation of the proposed method shows a reduction in computational complexity.

Methods

Selected machine learning algorithms, including linear regression, support vector regression (SVR) with and without nonlinear nuclear function, and evolutionary algorithms were used and compared in the modeling of glucose prediction. A description of each method is provided below.

Linear regression

When defining a linear model, the equation takes the general form:

$$\hat{Y} = bX + a$$

where:

predicted value of the dependent variable,

constant component of the equation,

regression coefficient

measured value of the independent variable

The regression coefficient is given by the formula (2)

$$b = \frac{n \cdot \sum x_i y_i - \sum x_i \cdot \sum y_i}{n \cdot \sum x_i^2 - \left(\sum x_i\right)^2}$$

where:

n-observation numbers

$\sum x_i$ -sum of all measurement values

$\sum y_i$ -sum of all measurement values

Support vector regression

For the regression analysis of the support vectors, which maps a non-linear set of multivariate data, the SVR methodology determines the prediction function, which is given by the formula (3):

$$f(x) = w^T \phi(x) + b$$

where:

trait space transformation,

weights of spatial matrices

random error

According to the Karush-Kuhn-Tucker theory [11], the regression function with the kernel function takes the form (4):

$$f(x) = \sum_{i=1}^{N} \left(a_i - a_i^*\right) k\left(x, x^i\right) + b$$

where:

Lagrange multipliers

kernel function used to compute the similarity between two input vectors in the transformed space

Modified evolutionary algorithm

Evolutionary algorithms are a global search strategy based on sample population analysis, where it is typically used as a global optimizer. Uses techniques in the analysis of differences in given samples (subjected to mutation, selection) on a multidimensional plane When minimizing a one-dimensional object, the objective function given by the formula (5) can be used:

$$F\left(x^*\right) = \min_{x, \in \Omega} f\left(x_i\right), x_i \in P = \prod_{j=1}^{D}\left[L_j, U_j\right]$$

where:

F-objective function

$x_i = \left(x_{i,1} \ldots x_{i,D}\right)$ -D-dimensional vector

x^* - global optimal solution to the optimization problem

L_j, U_j - limits of j-tej varible

P -search space,

The fitting function based on the evolutionary algorithm is given by the formula:

$$f(C, y, \varepsilon) = \frac{1}{n} \sum_{1}^{n} \left| \frac{R_i(C, y, \varepsilon) - F_i(C, y, \varepsilon)}{R_i(C, y, \varepsilon)} \right|$$

where:

control parameter,

kernel parameter,

current value,

predicted value,

 training set size

The optimal values of the parameters ()are found after the iterative operation of the algorithm is completed. For simulation studies, publicly available experimental data [12] were used as input data as the training set. These measurements were based on real-time sensor sampling (CGM) for continuous glucose monitoring. Forecasting was performed using machine learning algorithms. For each new sample taken, the model parameters are re-estimated. The mu factor represents the weight of the blood glucose sample in relation to the current sampling time.

Discussion of the results

The studies compared machine learning algorithms to predict blood glucose values to generate early warnings of hypoglycemia. An example of a blood glucose forecast is shown in Figure 1. The figure shows that, depending on the algorithms used, the predicted signal exceeds the hypoglycemic threshold in the range of 1-2. For linear regression, in the first case $t \approx 730$ minutes it is a false-positive result. In the second case, at $t \approx 830$ minutes, the model correctly predicts the event approximately 15 minutes in advance. The modified evolution algorithm and the SVR with nonlinear nuclear function are most effective in predicting glucose levels with an RMSE error below 8%. When using linear regression, the RMSE error varies between 32-11% depending on the dynamic change curves.
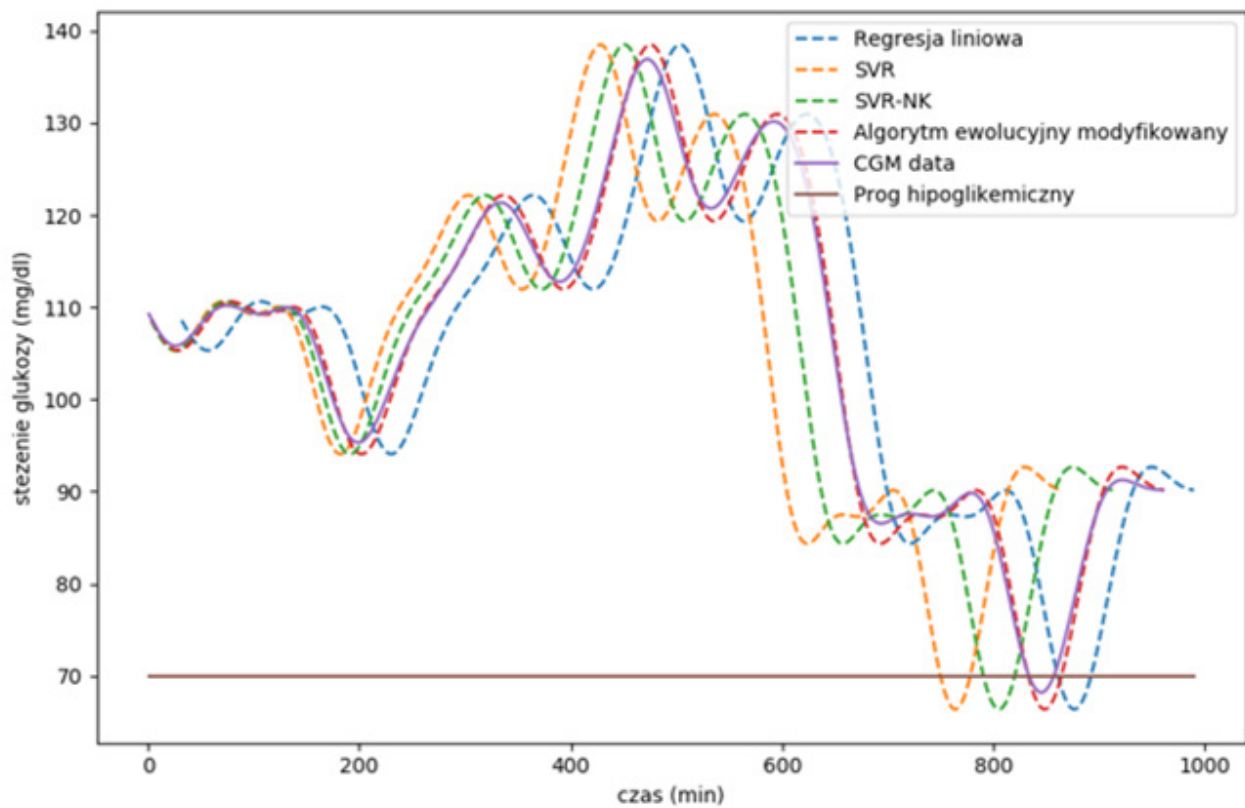

Figure 1: Results of blood glucose forecasting using various machine learning algorithms (ph=30 and mu = 0.9).

The proposed SVR technique similarly estimates non-linear changes in blood glucose while avoiding overfitting. Despite the rapid development of CGM systems as hardware devices, the evolution in the development of algorithms in blood glucose forecasting has been significantly less. The main problems of implementing devices for blood glucose monitoring systems are increasing the accuracy of predicting results, minimizing errors, as well as measuring accuracy of blood glucose concentration.

Summary

The paper presents selected machine learning algorithms for predicting blood glucose concentration. The support vector regression (SVR) algorithm and the evolutionary algorithm show the best prediction results. The evolutionary algorithm optimizes the SVR parameters and is most efficient. The proposed method has numerous advantages, such as: accuracy, adaptability, robustness, as well as efficiency in modeling nonlinear and complex data series. In the future, the clinical validation of the algorithm will continue for a potentially real application. Risk assessment is an adequate tool to predict the vulnerability of the system in terms of corrosion in order to manage a maintenance strategy for optimizing safety of human health.

References

1. Han J (2006) Accuracy and error measures concepts and techniques. Data mining, United States of America, p. 770.

2. Klonoff DC (2005) Continuous glucose monitoring-roadmap for 21st century diabetes therapy. Diabetes Care 28(5): 1231-1239.

3. Tan G, Yan J, Gao C, Suhua Y (2012) Prediction of water quality time series data based on least squares support vector machine. Procedia Engineering 31: 1194-1199.

4. Gertych A, Ing N, Ma Z, Fuchs TJ, Salman S, et al. (2015) Machine learning approaches to analyze histological images of tissues from radical prostatectomies. Comput Med Imaging Graph 46: 197-208.

5. Koutny T (2013) Glucose predictability, blood capillary permeability and glucose utilization rate in subcutaneous, skeletal muscle, and visceral fat tissues. Comput Biol Med 43: 1680-1686.

6. Bremer T, Gough D (1999) Is blood glucose predictable from previous values? A solicitation for data. Diabetes 48(3): 445-451.

7. Sparacino G, Zanderigo S, Aran A, Cobelli C (2006) Continuous glucose monitoring and hypo/hyperglycemia. Diabetes Res Clin Pract 74(2): 160-163.

8. Palerm C, Willis J, Desemone J, Bequette B (2005) Hypoglycemia prediction and detection using optimal estimation. Diabetes Technol Ther 7(1): 3-14.

9. Pappada SM, Cameron BD, Rosman PM (2008) Development of a neural network for prediction of glucose concentration in type 1 diabetes patients. J Diabetes Sci Technol 2(5): 792-801.

10. Pappada S, Cameron B, Rosman P, Bourey R, Papadimos T, et al. (2011) Neural network-based real-time prediction of glucose in patients with insulin dependent diabetes. Diabetes Technol Ther 13(2): 135-141.

11. Edgar TF, Himmelblau DM, Lasdon LS (2001) Optimization of chemical process. McGrawHill, New York, USA.

12. http://www.dei.unipd.it/~gianni/gsref-en.html

For possible submissions Click below:

Submit Article