Med Biol Eng Comput
DOI 10.1007/s11517-015-1320-9

SPECIAL ISSUE - ORIGINAL ARTICLE

Comparative assessment of glucose prediction models for patients with type 1 diabetes mellitus applying sensors for glucose and physical activity monitoring

K. Zarkogianni[1] · K. Mitsis[1] · E. Litsa[1] · M.-T. Arredondo[2] · G. Fico[2] · A. Fioravanti[2] · K. S. Nikita[1]

Received: 15 May 2014 / Accepted: 21 May 2015

Abstract The present work presents the comparative assessment of four glucose prediction models for patients with type 1 diabetes mellitus (T1DM) using data from sensors monitoring blood glucose concentration. The four models are based on a feedforward neural network (FNN), a self-organizing map (SOM), a neuro-fuzzy network with wavelets as activation functions (WFNN), and a linear regression model (LRM), respectively. For the development and evaluation of the models, data from 10 patients with T1DM for a 6-day observation period have been used. The models' predictive performance is evaluated considering a 30-, 60- and 120-min prediction horizon, using both mathematical and clinical criteria. Furthermore, the addition of input data from sensors monitoring physical activity is considered and its effect on the models' predictive performance is investigated. The continuous glucose-error grid analysis indicates that the models' predictive performance benefits mainly in the hypoglycemic range when additional information related to physical activity is fed into the models. The obtained results demonstrate the superiority of SOM over FNN, WFNN, and LRM with SOM leading to better predictive performance in terms of both mathematical and clinical evaluation criteria.

Keywords Self-organizing map · Neuro-fuzzy · Diabetes · Sensors · Physical activity · Glucose · Prediction · Neural networks

✉ K. Zarkogianni
kzarkog@biosim.ntua.gr

[1] Biomedical Simulations and Imaging Laboratory, National Technical University of Athens, 15780 Athens, Greece

[2] Technical University of Madrid, Madrid, Spain

1 Introduction

Type 1 diabetes mellitus (T1DM) is a chronic metabolic disease characterized by elevated blood glucose levels (BGLs) caused by the autoimmune destruction of the insulin-producing pancreatic beta cells resulting in the absence of insulin secretion. The subsequent lack of insulin leads to increased blood and urine glucose levels. The excess glucose circulating through the body in the blood stream over time leads to severe long-term-mortality-related complications such as cardiovascular disease, diabetic neuropathy, and diabetic retinopathy. T1DM can be controlled through exogenous insulin administration applying either multiple insulin injections or continuous subcutaneous insulin infusion through insulin pumps. However, the administration of too high insulin dose leads to hypoglycemic episodes. Intensive glycemic control, involving regular glucose measurements and exogenous insulin administration, is essential to reduce the occurrence of acute episodes such as severe hypoglycemia and hyperglycemia and long-term diabetes complications [22], and consequently increase patients' quality of life.

Continuous glucose measurement systems (CGMSs), able to provide with records of glucose levels every 1 min or 5 min, and insulin pumps are the latest technological advances within the scope of facilitating intensive glycemic control [11]. However, tight glycemic control is difficult to be achieved, since several environmental factors such as nutrition, physical activity, patient's psychological status and his overall lifestyle along with endogenous processes, such as circadian rhythms, strongly affect glucose metabolism. Furthermore, intra- and inter-patient variability in response to therapy makes the regulation of glucose levels very difficult. The aforementioned difficulties can be addressed through the development of computational models able to

Published online: 07 June 2015

produce accurate and reliable estimations of future glucose profile in response to various stimuli. Moreover, many attempts have been made in the recent years toward the development of closed-loop glucose controllers, which provide estimations of insulin infusion rates and boluses aiming at regulating glucose levels [16, 31]. These glucose controllers are usually based on model predictive control (MPC) in order to handle the delays and time lags caused by delivering insulin subcutaneously [9, 13, 16, 26, 31].

To this end, compartmental models (CMs) representing fundamental glucoregulatory processes have been developed [10, 13]. However, due to the fact that some of the endocrine processes affecting glucose metabolism are still not fully understood, these models take into account only a confined number of factors associated with glucose metabolism and cannot be easily individualized to accurately simulate metabolic processes for a specific T1DM patient. Moreover, the identification of CMs' parameters requires clinical measurements, which are not typically available in the clinical settings. In order to overcome these limitations, the use of data-driven modeling techniques has been proposed which disregard physiological insights and use pattern recognition techniques to simulate glucose metabolism. Volterra series models, time series analysis and machine learning methods are the most widespread data-driven techniques toward the development of glucose prediction models. In particular, nonlinear Volterra models of glucose–insulin dynamics have been shown to provide accurate predictions in the absence of noise [6, 14]. Autoregressive exogenous input (ARX) and Box–Jenkins (BJ) models with constant parameters and various model orders (high and low) have also been applied to simulate glucose–insulin dynamics [5]. Several types of artificial neural networks (ANNs) such as multilayer perceptron (MLP) neural networks (NN) [17, 18], radial basis function (RBF) NNs [2], wavelet NNs [28] and recurrent neural networks (RNNs) [2, 5, 6, 9, 10, 13–18, 26, 28, 31] have been deployed for the simulation of glucose metabolism. Furthermore, glucose prediction models based on Gaussian processes [23] have been developed. Additionally, hybrid glucose–insulin metabolism models based on the combined use of CMs and data-driven modeling techniques, such as RNN [15, 16, 31], SVR [8] and self-organizing maps (SOMs) [29], have produced prominent results. Most of these models are fed with information related to glucose records, insulin injections/infusion and meal intakes. Only a limited number of studies take into account physical activity as input to the models [7, 20, 24, 30].

The present work investigates the ability to produce accurate glucose profile predictions based on data provided by only two sensors, monitoring glucose concentration and physical activity, respectively. The potential of three models based on FNN, SOM, and a neuro-fuzzy network with wavelets as activation functions (WFNN), respectively, to capture the metabolic behavior of a specific patient with T1DM is investigated and comparatively assessed. The contribution of physical activity input data toward improvement of models prediction accuracy is evaluated. Issues related to low-input dimensionality, low complexity, stability and subsequent integration of the models in closed-loop glucose controllers, have motivated the investigation of these particular methodologies. SOM has the advantage of converting the input space into a low-dimensional map, which preserves the relations between input data. In the present study, the original SOM algorithm is modified in order to obtain multiple local linear models, making it, thus, suitable for subsequent integration into closed-loop glucose controllers. WFNN has the ability to handle low-input dimensionality but its subsequent integration into a MPC would result in high on line computational complexity caused by the direct use of nonlinear and non-convex programming techniques. FNN is chosen for comparison purposes, due to its widespread applications in the domain of dynamic modeling. Furthermore, aiming at justifying the need of applying such sophisticated techniques in order to capture the metabolic behavior of a T1DM patient, the performances of the three models have been compared with the one obtained by applying a linear regression model (LRM). For the development and evaluation of the models, data from 10 patients with T1DM, collected from a 6-day observation period, are used.

2 Materials and methods

2.1 Dataset

For the development and evaluation of the models, data collected within the framework of the European funded research project METABO [7] have been used. The data correspond to the medical records of 10 T1DM patients (7 males and 3 females) who were monitored for 10.70 ± 4.69 days.

During the observation period, the patients followed multiple-dose insulin therapy and wore the guardian real-time CGM system (Medtronic Minimed Inc.), which provided blood glucose records every 5 min, and the SenseWear Armband (BodyMedia Inc.) wearable body monitoring system, which recorded the energy expenditure of daily physical activities or exercise events with a resolution time of 1 min. In order to evaluate the models' ability to handle inter-subject variability, the data corresponding to identical number of days (6) for each patient were used. Table 1 summarizes the patients' characteristics.

Table 1 Patients' characteristics

	Average ± SD
Age	41.8 ± 14.39
Diabetes duration	25.00 ± 16.40
BMI	23.34 ± 3.49
HbA1c	7.75 ± 0.79

2.2 Methodology

Four machine learning methods based on FNN, SOM, WFNN, and LRM, respectively, are applied and comparatively assessed toward the development of personalized glucose prediction models. In order to assess the impact of the addition of physical activity data to the models' predictive performance, two different cases are investigated. In Case 1, two variables provide input to the models: (a) most recent glucose measurement ($G(t)$), and (b) glucose change ($\Delta G(t)$). In Case 2, information related to physical activity is added in the input space of Case 1. In particular, the sum of the energy expenditure during the time period [t—150 min, t—120 min] is fed into the models in order to take into account the physical activity during the latest 30 min with a lag time equal to 120 min. The lag time of 120 min is a constraint imposed by the requirement to evaluate the models' predictive performance up to 120 min PH.

2.2.1 Prediction models

2.2.1.1 Feedforward neural network One hidden layer, the hyperbolic tangent sigmoid transfer function and the backpropagation training algorithm have been used to develop the FNN.

2.2.1.2 Self-organizing map Although SOMs are usually applied for data clustering and visualization, they can also be successfully applied for dynamic modeling [29]. This is achieved by applying a vector quantization method [4]. In particular, a two-dimensional grid of N neurons is created and every neuron (i) is associated with a weight vector ($\boldsymbol{w_{in}}$) with dimension identical to that of the input vector and a weight output value (w_{out}) corresponding to the glucose prediction. During the training stage, the winning neuron is determined by calculating the Euclidean distance between the input vector and the weight vectors ($\mathbf{w_{in}}$). The neuron with the lowest Euclidean distance is the winner ($i^*(t)$),

$$i^*(t) = \arg\min\{\|\mathbf{x_{in}}(t) - \mathbf{w_{in}}(t)\|\} \tag{1}$$

where $\mathbf{x_{in}}$ denotes the input vector. The weights $\mathbf{w_{in}}$ and w_{out} are updated according to the following equations,

$$\Delta\mathbf{w_{in}}(t) = \lambda(t) \cdot h(i^*, i, t) \cdot [\mathbf{x_{in}}(t) - \mathbf{w_{in}}(t)] \tag{2}$$

$$\Delta w_{\text{out}}(t) = \lambda(t) \cdot h(i^*, i, t) \cdot [x_{\text{out}}(t) - w_{\text{out}}(t)] \tag{3}$$

where x_{out} is the actual glucose value, $\lambda(t)$ is the learning rate which decreases exponentially with time from value λ_0 until λ_T according to the equation,

$$\lambda(t) = \lambda_0 \cdot \left(\frac{\lambda_T}{\lambda_0}\right)^{\frac{t}{T}} \tag{4}$$

where T is the number of total epochs. Furthermore, $h(i^*, i, t)$ is the neighborhood function with a Gaussian form,

$$h(i^*, i, t) = \mathrm{e}^{\left(\frac{-\|r_i(t) - r_{i^*}(t)\|^2}{2 \cdot s(t)^2}\right)} \tag{5}$$

where $r_i(t)$ and $r_{i^*}(t)$ are the locations of neurons i^* and i, respectively, while $s(t)$ is the variance of the Gaussian function and decreases exponentially with time from value s_0 until s_T,

$$s(t) = s_0 \cdot \left(\frac{s_T}{s_0}\right)^{\frac{t}{T}} \tag{6}$$

Thus, at the beginning of the training procedure, a wide area of neurons is affected and as the procedure goes by, a smaller area around the winner neuron is affected until only the winner neuron changes its weights.

After the training stage, SOM can be used to obtain estimates of future glucose values. For every new input vector, the winner neuron (i^*) is calculated and the corresponding w_{out} value is the glucose prediction. However, this technique requires a great number of neurons in order to have a small prediction error [4]. To overcome this problem, a technique that allows the creation of multiple local linear models is applied [19]. Particularly, the estimated glucose levels are produced from a linear autoregressive model with exogenous inputs, as follows,

$$u_{\text{pred}}(t+1) = \mathbf{v}^{\mathrm{T}} \cdot \mathbf{x}_{\text{in}} \tag{7}$$

where $\mathbf{v}$ represents linear models' coefficients which are calculated during the training stage [3, 25] starting from zero values. The vector of these coefficients ($\mathbf{v}_i$) corresponding to each neuron is updated according to the equation,

$$\Delta\mathbf{v}_i(t) = \lambda(t) \cdot h(i^*, i, t) \cdot \left[x_{\text{out}}(t+1) - \mathbf{v}_i(t)^{\mathrm{T}} \cdot \mathbf{x_{in}}(t)\right] \cdot \frac{\mathbf{x_{in}}(t)}{\|\mathbf{x_{in}}(t)\|^2} \tag{8}$$

Thus, the neighborhood around every neuron forms a local linear model. After the training stage, when a new input vector arises, the winner neuron is computed and the corresponding coefficients are used to produce the future glucose value using Eq. (7).

Parameter N and number of epochs are determined through trial and error, which is performed by applying the 10-fold cross-validation for each individual patient. In particular, several candidate values for the parameter N and the number of epochs are tested and the optimum values achieving the lowest average validation root-mean-square error (RMSE) calculated over all folds over all patients are found. A grid of 15 $\times$ 15 is found to be large enough to quantize the input data. Moreover, a number of 50 epochs is sufficient to achieve convergence. Regarding the parameters associated with the learning rate, λ_0 and λ_T are set equal to 0.9 and 0.01, respectively, s_0 is equal to 14 which corresponds to the maximum distance between all neurons, while s_T is set equal to 1. The initial values of the weights are randomly selected within the range [0, 1]. It is also noted that the data have been normalized within the range [0, 1].

2.2.1.3 Neuro-fuzzy network with wavelets as activation functions Neuro-fuzzy-based systems take advantage of the fuzzy logic's ability to use limited information and provide solutions to problems that are not well structured and the NNs' ability to be trained and generalized. In particular, the fuzzy sets and rules are approximated through NNs since limited empirical knowledge exists to create the appropriate IF–THEN rules [1]. Furthermore, in order to avoid trapping into local minima and slow convergence during training stage, wavelets are applied as activation functions. The overall architecture of the WFNN consists of seven layers and is presented in Fig. 1. The first layer corresponds to the input layer. In the second layer, the fuzzification process is applied, using Gaussian functions as membership functions of the fuzzy logic. It should be noted that the choice of the Gaussian functions is based on the need of

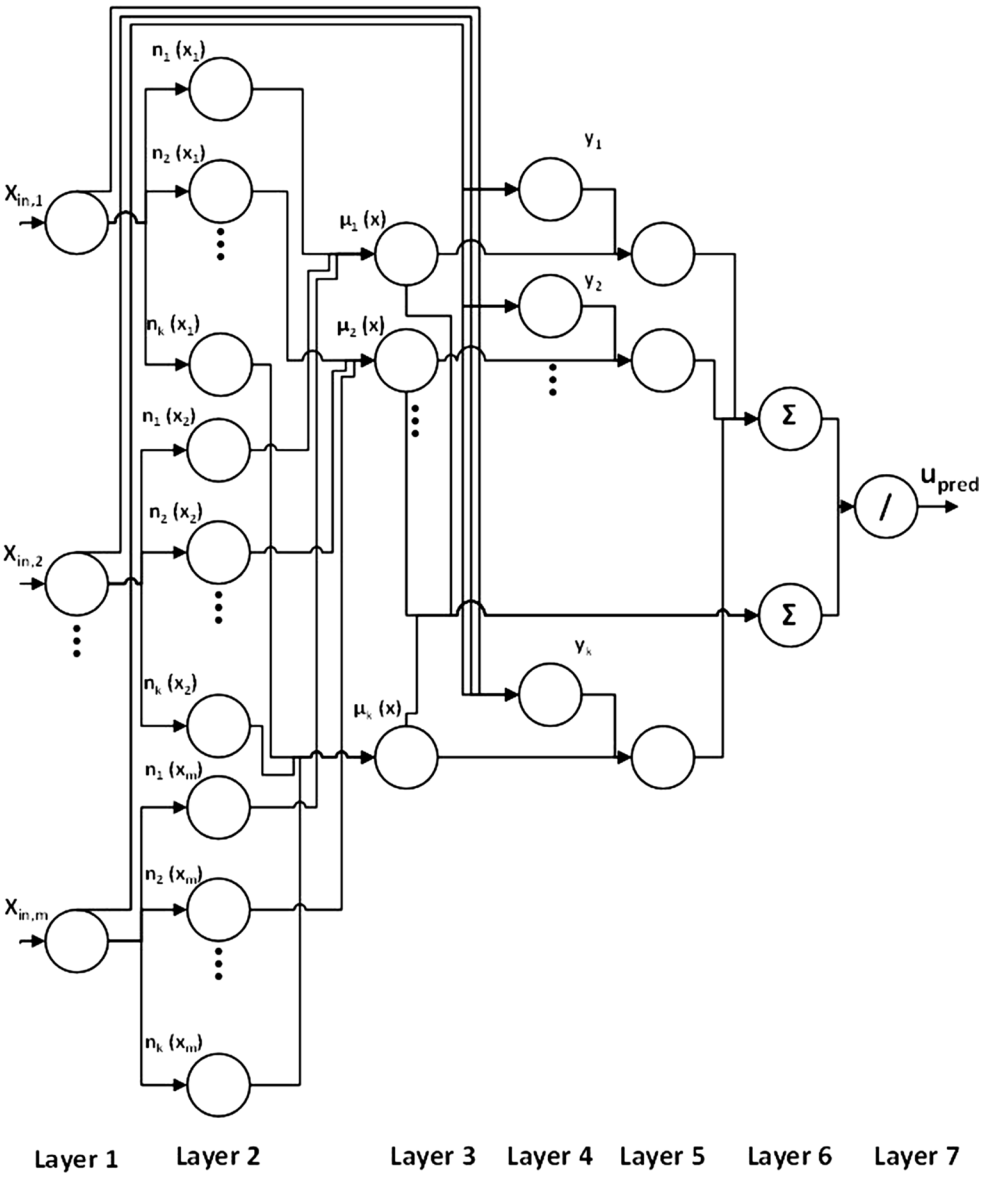

Fig. 1 Overall architecture of WFNN

obtaining a degree of belonging to a given interval and due to its smoother transitions compared to trapezoidal function. Each input is associated with k membership functions,

$$n_j(x_{\text{in},i}) = \mathrm{e}^{-\frac{(x_{\text{in},i}-c_{ij})^2}{\sigma_{ij}^2}}, \quad \begin{matrix} j=1,\ldots,k \\ i=1,\ldots,m \end{matrix} \tag{9}$$

where $n_j(x_{\text{in},i})$ is the membership degree of the ith input in the jth fuzzy set, k is the number of membership functions, and m is the number of inputs, while c_{ij} and σ_{ij} represent the centers and the standard deviations of the Gaussian functions, respectively. In the third layer, the fuzzy rules (μ) are generated. Each membership function participates in only one rule. The outcome of the rules is calculated using the minimum operation, as follows,

$$\mu_j(x_{\text{in}}) = \prod_i n_j\left(x_{\text{in},i}\right), \quad \begin{matrix} j=1,\ldots,k \\ i=1,\ldots,m \end{matrix} \tag{10}$$

In the fourth layer, the wavelet activation functions (ψ) are calculated. The wavelet function used is the mexican hat,

$$\begin{aligned} \psi(z_{ij}) &= \frac{1}{\sqrt{|a_{ij}|}}\left(1-z_{ij}^2\right)\cdot \mathrm{e}^{\frac{-z_{ij}}{2}}, \quad \begin{matrix} j=1,\ldots,k \\ i=1,\ldots,k \end{matrix} \\ z_{ij} &= \frac{x_{\text{in},i}-b_{ij}}{a_{ij}} \end{aligned} \tag{11}$$

where a_{ij}, b_{ij} are the dilation and translation of the wavelet function, respectively. The output of the fourth layer is calculated as

$$y_l = w_l \cdot \Psi_l(z), \quad l = l,\ldots,k \tag{12}$$

where

$$\Psi_l(z) = \sum_{i=1}^{m} \psi(z_{ij}) \tag{13}$$

The defuzzification process is carried out in the fifth and sixth layer. Particularly, the outputs of the third and fourth layer are multiplied and then divided by the sum of all the outputs of the third layer. The output is generated in the seventh layer as

$$u_{\text{pred}} = \frac{\sum_{l=1}^{k} \mu_l(x_{\text{in}})\cdot y_l}{\sum_{l=1}^{k} \mu_l(x_{\text{in}})} \tag{14}$$

During the training stage, the parameters of the Gaussian functions (c_{ij}, σ_{ij}) along with the weights (w_l) and the parameters of the wavelets (a_{ij}, b_{ij}) are updated according to a gradient-based algorithm with adaptive learning rate. In particular, if the RMSE of a certain epoch is greater than that of the previous epoch, the learning rate is reduced slightly. Similarly, if the RMSE is smaller, the learning rate slightly increases. It should be stressed that in this work the mini-batch method [27] is used, which is a hybrid method combining the batch and online method for training NNs.

The number of the membership functions (k) along with the number of epochs are defined through trial and error and set equal to 6500, respectively. In particular, these are the optimum values obtained by applying the 10-fold cross-validation individually for each patient in order to test several candidate values and correspond to the ones achieving the lowest average validation RMSE calculated over all folds over all patients. The initial value of the learning rate is chosen to be 0.01, while taking into consideration that a large learning rate can lead to unstable learning and a very small one slows the algorithms learning speed significantly. The number of subsets in which the original trained set is divided in order to apply the hybrid method of training is 40, which is defined, taking into account the size of the training sets (approximately 1100). Furthermore, the data have been normalized within the range [0, 1].

2.2.1.4 Linear regression model The LRM has the following form

$$u_{\text{pred}} = \boldsymbol{\beta}^{\tau} \cdot \boldsymbol{x_{in}} \tag{14}$$

where $\boldsymbol{\beta}$ represents the vector of coefficients, which are calculated by applying the least-squares method.

2.2.2 Evaluation criteria

In order to account for the inter-subject variability, the models are patient specific by being trained individually on each patient's data. Aiming at assessing the models' generalization ability, the 10-fold cross-validation criterion is applied individually for each patient. The folds are generated randomly by applying the MATLAB function "crossvalid." It should be stressed that in order to achieve a fair comparison of the models' predictive performance, the folds created for each patient remain the same when applying the different models. The predictive performance of the models is evaluated considering PHs of 30, 60 and 120 min with a 5-min resolution, using both mathematical and clinical evaluation criteria. In particular, the RMSE, correlation coefficient (CC) and the mean absolute relative difference (MARD) are applied to evaluate the performance of the models in terms of matching the predicted glucose profiles with the original ones. Furthermore, the continuous glucose-error grid analysis (CG-EGA) [12] is applied to evaluate the clinical accuracy of the glucose predictions and their effect on decisions to avoid hypo- and hyper-glycemic events.

3 Results

3.1 Evaluation of models' predictive performance in Case 1

Mathematical and clinical evaluation criteria, as described above, are applied to evaluate the models' predictive performance, in case of being fed with information related to glucose and glucose change. In particular, Table 2 presents the average and standard deviation values of RMSE, CC and MARD over all patients which are obtained by applying the three glucose prediction models considering PHs of 30, 60, and 120 min. In order to investigate whether there is at least one significant difference between

Table 2 Mathematical criteria for evaluating the models' predictive performances in Case 1

PH (min)	Model	RMSE	CC (%)	MARD
30	FNN	17.81 ± 8.80	95.42 ± 3.88	7.25 ± 1.48
	SOM	**12.29 ± 2.27**	**97.92 ± 0.70**	**5.34 ± 1.08**
	WFNN	15.64 ± 3.15	96.87 ± 0.87	7.38 ± 1.77
	LRM	15.51 ± 3.78	96.81 ± 0.72	7.47 ± 2.08
60	FNN	25.70 ± 9.23	90.80 ± 4.62	11.44 ± 2.54
	SOM	**21.06 ± 3.20**	**94.00 ± 1.77**	**9.36 ± 1.95**
	WFNN	25.5 ± 4.62	91.72 ± 2.18	12.38 ± 2.83
	LRM	26.39 ± 7.98	91.22 ± 2.10	12.79 ± 4.17
120	FNN	38.51 ± 9.67	79.12 ± 5.26	18.68 ± 5.26
	SOM	**33.68 ± 5.26**	**84.22 ± 4.87**	**15.99 ± 3.14**
	WFNN	40.81 ± 7.09	79.06 ± 5.26	20.42 ± 4.55
	LRM	44.52 ± 19.31	77.99 ± 5.40	22.27 ± 9.62
Average ± SD (over all PHs)	FNN	27.34 ± 10.44	88.44 ± 8.40	12.45 ± 5.78
	SOM	22.34 ± **10.75**	92.04 ± **7.05**	10.23 ± **5.37**
	WFNN	27.31 ± 12.68	89.21 ± 9.16	13.39 ± 6.57
	LRM	28.80 ± 14.65	88.67 ± 9.66	14.17 ± 7.49

The best values obtained for each criterion are shown in bold

Table 3 *P*-values obtained by applying *t*-test to compare the models' predictive performance in Case 1

	PH 30 min			PH 60 min			PH 120 min		
	WFNN	SOM	LRM	WFNN	SOM	LRM	WFNN	SOM	LRM
RMSE									
FF	0.19	**0.03**	0.22	0.47	**0.04**	0.42	0.13	0.02	0.14
WFNN		**0.00**	0.42		**0.00**	0.29		**0.00**	0.22
SOM			**0.00**			**0.01**			**0.03**
CC									
FF	0.14	**0.04**	0.16	0.28	**0.03**	0.40	0.49	**0.01**	0.30
WFNN		**0.00**	0.37		**0.00**	0.11		**0.00**	0.15
SOM			**0.00**			**0.00**			**0.00**
MARD									
FF	0.33	**0.00**	0.31	**0.03**	**0.00**	0.06	**0.02**	**0.01**	0.06
WFNN		**0.00**	0.42		**0.00**	0.32		**0.00**	0.22
SOM			**0.00**			**0.01**			**0.02**
AR (Hypo)									
FF	0.23	**0.00**	0.08	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**
WFNN		**0.00**	0.08		**0.00**	0.14		0.13	0.40
SOM			**0.00**			**0.02**			0.09
AR (Hyper)									
FF	**0.01**	**0.00**	0.10	0.09	**0.00**	0.07	0.16	**0.00**	0.09
WFNN		**0.00**	**0.00**		**0.00**	**0.01**		**0.00**	0.09
SOM			**0.00**			**0.00**			**0.01**

the models' predictive performance, the one-way analysis of variance (ANOVA) has been applied with a 0.05 level of significance. The ANOVA revealed statistical significant differences in terms of CC (p value 0.04 for 120 min PH) and MARD (p value 0.02, for 30 min PH). In order to perform a pairwise comparison of the models' accuracy, the one-tail paired t test has been applied and the cutoff 0.05 has been considered as the level of significance. In Table 3, the corresponding p values are presented. It can be shown that SOM is statistically more accurate than FNN, WFNN and LRM in terms of RMSE, CC and MARD for all PHs. FNN, LRM and WFNN achieve similar accuracy for 30 min PH, while for greater PHs, statistical significant differences are observed between WFNN and FNN in terms of MARD. The CG-EGA [12] is shown in Fig. 2, where the percentage of accurate readings (AR), benign errors (BE), and erroneous errors (ER) in each relevant glucose range (hypoglycemia, euglycemia and hyperglycemia) are depicted. According to the CG-EGA, each model demonstrates its lowest performance in the hypoglycemic range, while most ARs have been achieved by applying the SOM (86.06, 72.53, 54.74 % for PH up to 30, 60, and 120 min, respectively). In the euglycemic and hyperglycemic range, all the three models perform well even in the case of PH up to 120 min. The ANOVA and the paired t test are applied in order to compare the ARs in the hypoglycemic and hyperglycemic range. Significant differences in terms of ARs in both the hypoglycemic

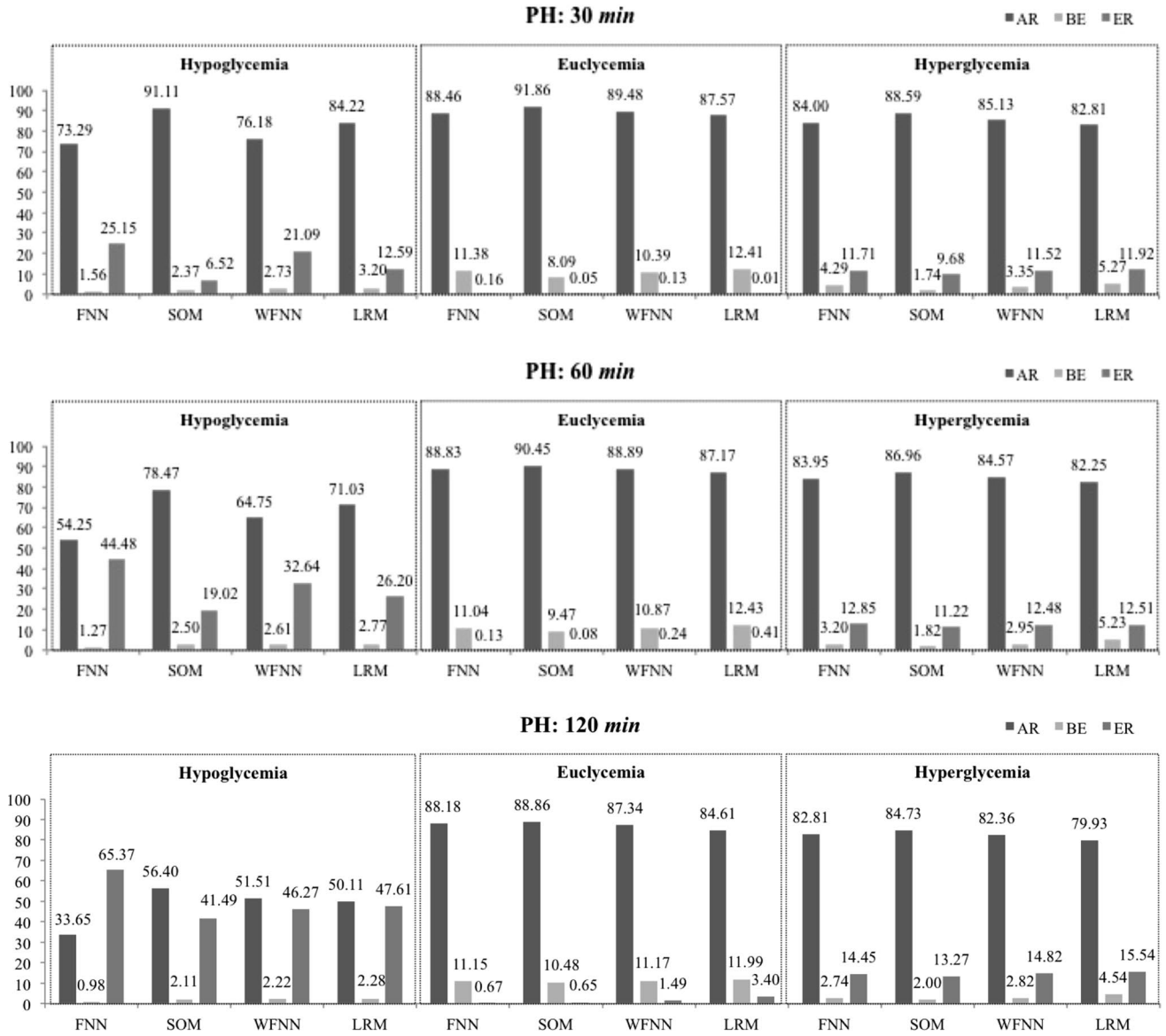

Fig. 2 CG-EGA applied in Case 1 to evaluate the glucose predictions' clinical accuracy

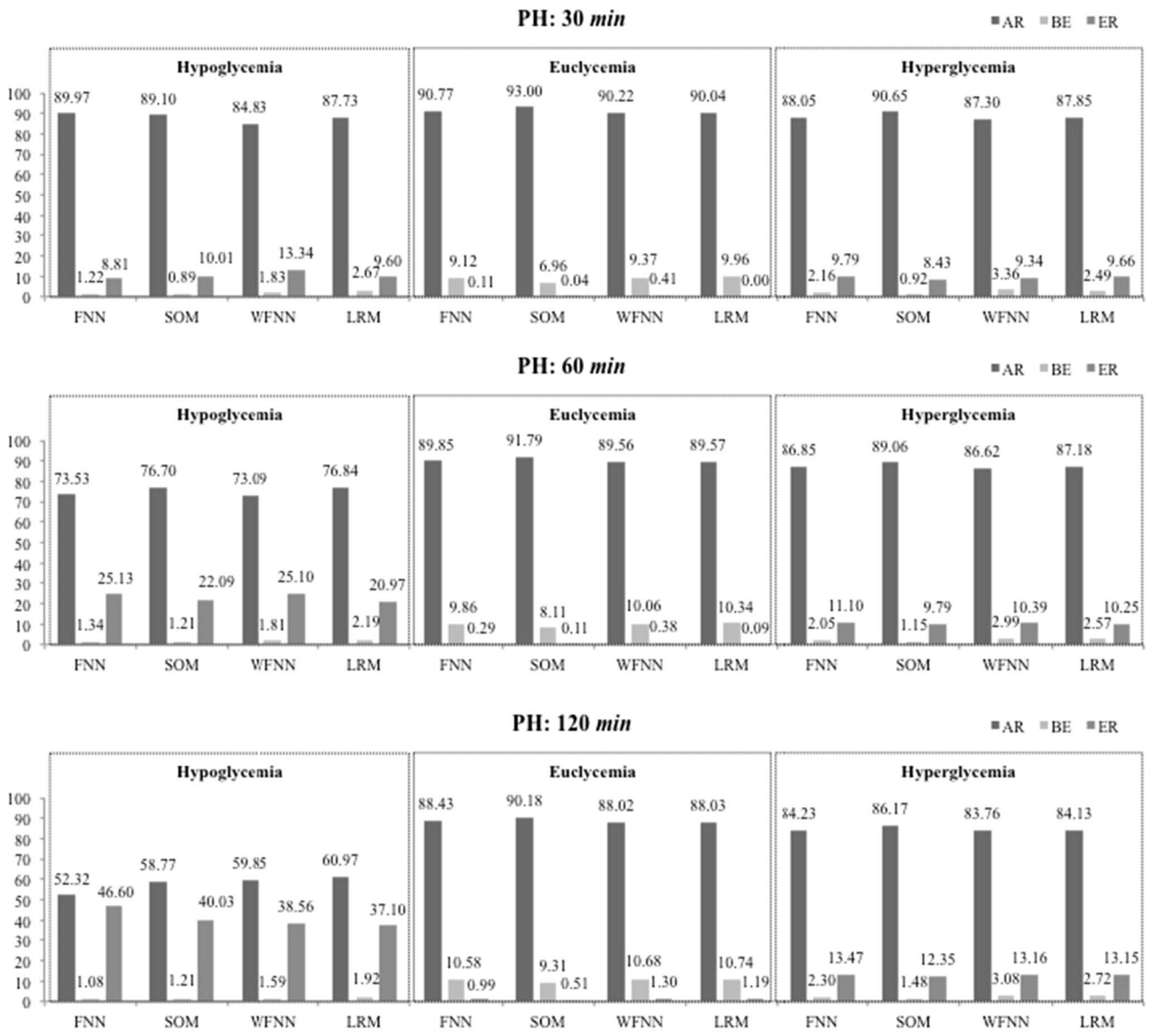

Fig. 3 CG-EGA applied in Case 2 to evaluate the glucose predictions' clinical accuracy

(p value 0.00 for all PHs) and the hyperglycemic range (p value 0.03 for 30 min PH) have been observed from applying the one-way ANOVA. The t test has revealed SOM's superior performance in the hypo- and hyper-glycemic range over the other models for all PHs (Table 3). The models' predictive performance deteriorates for larger PHs, as demonstrated from the mathematical criteria presented in Table 2. Furthermore, it is concluded that SOM achieves the smallest variations in RMSE, CC and MARD with respect to the different PHs as indicated from the standard deviations (SDs) over all PHs which are also presented in Table 2. Thus, SOM appears to be the most robust model for larger PHs.

3.2 Evaluation of models' predictive performance in Case 2

The results obtained by feeding the models with information related to glucose, glucose change and physical activity are shown in Tables 4 and 5 and Fig. 3. RMSE, CC and MARD average and standard deviation values over all patients are summarized in Table 4. It is observed that the additional input related to physical activity results in improved predictive performance. In particular, RMSE value as averaged over all PHs is reduced by a factor of 10, 7 and 3 % for FNN, SOM and WFNN, respectively. Similarly, MARD value as

Table 4 Mathematical criteria for evaluating the models' predictive performances in Case 2

PH (min)	Model	RMSE	CC (%)	MARD
30	FNN	13.31 ± 4.47	97.34 ± 1.54	5.65 ± 1.76
	SOM	**11.42 ± 2.33**	**98.14 ± 0.37**	**5.19 ± 1.48**
	WFNN	15.22 ± 2.17	96.73 ± 0.77	7.28 ± 1.52
	LRM	13.88 ± 2.55	97.14 ± 0.66	6.60 ± 1.56
60	FNN	22.66 ± 6.86	92.37 ± 3.53	10.07 ± 3.12
	SOM	**19.58 ± 3.80**	**94.26 ± 1.27**	**8.95 ± 2.24**
	WFNN	24.66 ± 3.39	91.56 ± 1.91	11.74 ± 2.47
	LRM	23.65 ± 3.93	91.88 ± 1.86	11.26 ± 2.69
120	FNN	37.62 ± 11.79	79.81 ± 7.70	17.28 ± 5.40
	SOM	**31.00 ± 6.07**	**84.28 ± 6.54**	**14.56 ± 3.46**
	WFNN	39.59 ± 5.03	78.89 ± 5.20	18.83 ± 3.83
	LRM	39.20 ± 6.49	79.19 ± 5.57	19.17 ± 4.81
Average ± SD (over all PHs)	FNN	24.53 ± 12.26	89.84 ± 9.03	11.00 ± 5.87
	SOM	20.67 ± **9.84**	92.23 ± **7.15**	9.57 ± **4.72**
	WFNN	26.49 ± 12.29	89.06 ± 9.18	12.62 ± 5.82
	LRM	25.57 ± 12.79	89.40 ± 9.22	12.34 ± 6.35

Table 5 *P* values obtained by applying *t* test to compare the models' predictive performance in Case 2

	PH 30 min			PH 60 min			PH 120 min		
	WFNN	SOM	LRM	WFNN	SOM	LRM	WFNN	SOM	LRM
RMSE									
FF	0.09	**0.04**	0.31	0.17	**0.03**	0.28	0.30	**0.02**	0.31
WFNN		**0.00**	0.04		**0.00**	0.19		**0.00**	0.40
SOM			**0.00**			**0.00**			**0.00**
CC									
FF	0.16	0.07	0.34	0.28	0.07	0.32	0.39	0.07	0.38
WFNN		**0.00**	0.15		**0.00**	0.38		**0.02**	0.45
SOM			**0.00**			**0.00**			**0.00**
MARD									
FF	**0.01**	0.07	**0.00**	0.08	**0.03**	**0.02**	0.22	**0.01**	**0.04**
WFNN		**0.00**	0.17		**0.01**	0.34		**0.02**	0.43
SOM			**0.00**			**0.00**			**0.00**
AR (Hypo)									
FF	0.07	0.26	0.25	0.47	0.16	0.31	0.17	**0.03**	0.22
WFNN		0.11	0.23		0.24	0.25		0.44	0.45
SOM			0.26			0.47			0.31
AR (Hyper)									
FF	0.34	**0.00**	0.38	0.43	**0.00**	0.33	0.38	**0.00**	0.45
WFNN		0.06	0.36		0.09	0.34		0.08	0.38
SOM			**0.00**			**0.01**			**0.01**

averaged over all PHs is reduced by a factor of 11, 6, and 5 % for FNN, SOM and WFNN, respectively. CC value as averaged over all PHs is increased by a factor of 1 and 0.2 % for SOM and FNN, respectively, while it is slightly reduced (0.1 %) for WFNN. Figure 3 presents the results of CG-EGA in Case 2. Comparing the results of CG-EGA depicted in Fig. 3, with the ones obtained in Case 1, it can be observed that the hypoglycemic range benefits the most from the additional physical activity information, followed by the hyperglycemic range. In the euglycemic range, a rather small increase in ARs is observed (Fig. 3).

Similarly to Case 1, SOM achieves the lowest RMSE and MARD values and the highest CC value as compared to FNN, WFNN and LRM. The application of the one-way ANOVA has revealed statistical significant differences between the models' predictive performance in terms of CC (p value 0.02 for 30 min PH) and MARD (p value 0.02 for 30 min PH). Table 5 summarizes the p values obtained by applying the paired t test. Statistical significant differences are observed between SOM and the other models in terms of RMSE, CC and MARD, demonstrating, thus, SOM's best predictive performance. Regarding the models' performance in the hypoglycemic range of the CG-EGA, according to the results obtained by applying the t test in order to compare the average ARs (Table 5), all the models produce similar results for PHs up to 60 min, while SOM outperforms FNN for PH equal to 120 min. Regarding the hyperglycemic range, SOM outperforms FNN and LRM, while SOM and WFNN demonstrate similar performance for all PHs. In the euglycemic, all the models perform similarly and well for all PHs.

4 Discussion

In this work, four machine learning methodologies based on FNN, SOM, WFNN and LRM toward the development of glucose prediction models have been applied and comparatively assessed. The models are fed with data from sensors monitoring glucose and physical activity. The input space consists of the most recent glucose measurement and glucose change, along with or without information related to energy expenditure. The objective of this work is to compare the four models in terms of their ability to accurately predict glucose profiles, and handle inter- and intra-patient variability based only on most recent information of glucose and physical activity. Low-input dimension provides the advantage of keeping the time necessary to train the models relatively small, which is important especially for the WFNN model since, in general, neuro-fuzzy systems require large training time for large-dimensional input spaces.

Although all the models perform well in the euglycemic range, the most accurate glucose predictions in the hypoglycemic and hyperglycemic range are produced by the SOM-based model. This, along with the fact that SOM results in better RMSE, CC and MARD values, proves SOM's superiority over FNN, WFNN and LRM for both cases of input data. It should be stressed, though, that the greater number of SOM's training parameters compared to the number of the other models' training parameters gives advantage to the SOM and might increase the risk of overfitting, especially considering the fact that the number of training instances is not enough higher than the SOM's training parameters. However, the application of the 10-fold cross-validation, which gives an insight on how the model generalizes to an independent dataset, limits the problem of overfitting. It is also noteworthy that the introduction of physical activity to the models leads to more accurate glucose predictions especially in the hypoglycemic range. However, it should be noted that the WFNN is little affected by the inclusion of physical activity data in the input space, demonstrating, thus, its ability to perform well even in the case of receiving less informative input.

A comparison between the results obtained by applying the SOM model and those reported in the literature is carried out. Although direct and fair comparison is not feasible due to different datasets and input spaces, substantial inferences can be obtained. Previous relevant studies focusing on the development of glucose prediction models for patients with T1DM based on methods different from the ones applied in the present study are chosen for comparison purposes. Furthermore, taking into consideration that the input space usually includes information related to either glucose levels or glucose levels along with insulin injections/infusion rates, meal intakes and physical activity, relevant studies referring to the former case are chosen for comparison purposes. In particular, the application of autoregressive models with time-varying parameters for the estimation of future glucose profile taking as input past CGM data [21], resulted in RMSE equal to 18.78 mg/dl for PH up to 30 min which is higher than the one obtained by applying SOM (12.29 mg/dl) and WFNN (15.64 mg/dl) in Case 1. Moreover, a glucose prediction model based on SVR, which received as inputs past CGM data, achieved 15.29, 24.19 and 33.04 mg/dl RMSE values for PHs up to 30, 60 and 120 min, respectively [8]. These values are higher than the ones obtained by SOM (12.29, 21.06, 33.68 mg/dl for 30, 60 and 120 min PH, respectively) and smaller than those obtained by WFNN (15.64, 25.50, 40.81 mg/dl for 30, 60 and 120 min PHs, respectively) in Case 1. A glucose predictor implemented with a multilayer FNN [18] resulted in around 18 mg/dl RMSE value for 30 min PH versus 12.29 and 15.64 mg/dl obtained by SOM and WFNN, respectively. From the above, SOM's superiority over the other models is demonstrated. The advantage of SOM relies on its ability to provide clinically acceptable glucose predictions through the creation of localized linear models. Moreover, taking into account that, the SOM is based on quantizing the input space, generalization can be ensured, if for every new input, a similar neuron exists. This does not necessarily require large number of training instances, but the training data should cover adequately the range of glucose values (in the hypoglycemic, euglycemic and hyperglycemic range) and glucose changes as a result of the patient's lifestyle and treatment. Future research directions refer to further improve the SOM-based model by choosing patients' specific number of neurons and epochs.

5 Conclusions

Four glucose prediction models based on FNN, SOM, WFNN and LRM have been developed, evaluated and comparatively assessed. The models are fed with data originating from wearable sensors monitoring glucose. Furthermore, the impact of physical activity input data to models' predictive performance has been investigated. For the development and evaluation of the models, data from 10 patients with T1DM have been used. The superiority of SOM over FNN, WFNN and LRM has been demonstrated. The obtained results indicate the SOM's ability to capture the metabolic behavior of a patient with T1DM and to handle inter- and intra-patient variability. SOM's low complexity and the fact that it results in linear glucose prediction models, make it a good candidate for integration into a closed-loop glucose controller.

References

1. Abiyev R, Kaynak O (2008) Fuzzy wavelet neural networks for identification and control of dynamic plants—a novel structure and comparative study. IEEE Trans Ind Electron 55(8):1–4
2. Baghdadi G, Nasrabadi AM (2007) Controlling blood glucose levels in diabetics by neural network predictor. In: Proceedings of the 29th annual international conference of the IEEE engineering in medicine and biology society, pp 3216–3219, Aug 22–26
3. Barreto G (2007) Time series prediction with the self-organizing map: a review. In: Hammer B, Hitzler p (eds) Perspectives of neural-symbolic integration, vol 77. Springer, Berlin, pp 135–158
4. Barreto G, Araujo A (2004) Identification and control of dynamical systems using the self-organizing map. IEEE Trans Neural Netw 15(5):1244–1259
5. Finan DA, et al (2006) Identification of linear dynamic models for type 1 diabetes: a simulation study. In: Proceedings of 14th IFAC ADCHEM symposium, pp 503–508, Apr 2–5
6. Florian JA, Parker RS (2005) Empirical modeling for glucose control in diabetes and critical care. Eur J Control 11:344–350
7. Georga E, et al (2009) Data mining for blood glucose prediction and knowledge discovery in diabetic patients: the METABO diabetes modeling and management system. In: Proceedings of the 31st annual international conference IEEE engineering in medicine and biology society, Minneapolis, 2009
8. Georga E et al (2013) Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression. IEEE Biomed Health Inform 17:71–81
9. Gillis R, Palerm CC, Doyle FJ (2007) Glucose estimation and prediction through meal responses using ambulatory subject data for advisory mode model predictive control. J Diabetes Sci Technol 1:825–833
10. Hovorka R et al (2004) Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. Physiol Meas 25:905–920
11. Klonoff DC (2005) Continuous glucose monitoring: roadmap for 21st century diabetes therapy. Diabetes Care 28(5):1231–1239
12. Kovatchev B et al (2004) Evaluating the accuracy of continuous glucose-monitoring sensors. Diabetes Care 27(8):1922–1928
13. Magni L et al (2009) Model predictive control of glucose concentration in type 1 diabetic patient: an in silico trial. Biomed Signal Process Control 4:338–346
14. Mitsis GD, Markakis MG, Marmarelis VZ (2009) Nonlinear modeling of the dynamic effects of infused insulin on glucose: comparison of compartmental with Volterra models. IEEE Trans Biomed Eng 56:2347–2358
15. Mougiakakou SG et al (2008) Prediction of glucose profile in children with type 1 diabetes mellitus using continuous glucose monitors and insulin pumps. Horm Res 70:22–23
16. Mougiakakou SG et al (2010) SMARTDIAB: a communication and information technology approach for the intelligent monitoring, management and follow-up of type 1 diabetes patients. IEEE Trans Inf Technol Biomed 14(3):622–633
17. 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 2:792–801
18. Perez-Gandia C et al (2010) Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring. Diabetes Technol Ther 12(1):81–88
19. Principe J, Wang L, Motter M (1998) Local dynamic modeling with self-organizing maps and applications to nonlinear system identification and control. Proc IEEE 86(11):2240–2258
20. Rollins D et al (2010) Free-living inferential modelling of blood glucose level using only noninvasive inputs. J Process Control 20:95–107
21. Sparacino G et al (2007) Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. IEEE Trans Biomed Eng 54(5):931–937
22. The DCCT Research Group (1993) The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin dependent diabetes mellitus. N Engl J Med 329(14):977–986
23. Valleta J, Chipperfield A, Byrne C (2009) Gaussian process modelling of blood glucose response to free-living physical activity data in people with type 1 diabetes. In: Proceedings of the 1st annual international conference of the IEEE engineering in medicine and biology society, pp 4913–4916, Sept 3–6
24. Valleta J, Chipperfield A, Byrne C (2009) Gaussian process modelling of blood glucose response to free-living physical activity data in people with type 1 diabetes. In: Proceedings of the 1st annual international conference of the ieee engineering in medicine and biology society, pp 4913–4916
25. Walter J, Ritter H, Schulten K (1990) Non-linear prediction with self-organizing maps. In: Proceedings of the IEEE international joint conference on neural networks (IJCNN'90), vol 1, pp 589–594, June 17–21
26. Wang Y, Dassau E, Doyle FJ (2010) Closed-loop control of artificial pancreatic β-cell in type 1 diabetes mellitus using model predictive iterative learning control. IEEE Trans Biomed Eng 57(2):211–219
27. Wilson D, Martinez T (2003) The general inefficiency of batch training for gradient descent learning. Neural Netw 16:1429–1451
28. Zainuddin Z, Pauline O, Ardil C (2009) A neural network approach in predicting the blood glucose level for diabetic patients. Int J Comput Intell 5(1):72–79
29. Zarkogianni K, et al (2013) Personalized glucose–insulin metabolism model based on self-organizing maps for patients with type 1 diabetes mellitus. In 13th IEEE international conference on bioinformatics and bioengineering (BIBE 2013), pp 1–4, Nov 10–13
30. Zarkogianni K, et al (2014) Neuro-fuzzy based glucose prediction model for patients with type 1 diabetes mellitus. In: IEEE-EMBS international conferences on biomedical and health informatics (BHI), June 1–4
31. Zarkogianni K et al (2011) An insulin infusion advisory system based on autotuning nonlinear model-predictive control. IEEE Trans Biomed Eng 58(9):2467–2477