Applied Soft Computing 134 (2023) 110012

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Reducing high-risk glucose forecasting errors by evolving interpretable models for Type 1 diabetes

A. Della Cioppa [a,b,*], I. De Falco [b], T. Koutny [c,d], U. Scafuri [b], M. Ubl [c], E. Tarantino [b]

[a] NCLab, DIEM - University of Salerno, Fisciano (SA), 84084, Italy
[b] Institute of High Performance Computing and Networking, National Research Council of Italy, Naples, 80131, Italy
[c] DCSE – Department of Computer Science and Engineering, University of West Bohemia, Pilsen 306 14, Czech Republic
[d] NTIS – New Technologies for the Information Society, University of West Bohemia, Pilsen 306 14, Czech Republic

ARTICLE INFO

Article history:
Received 26 July 2022
Received in revised form 22 November 2022
Accepted 6 January 2023
Available online 11 January 2023

Keywords:
Glucose forecasting
Evaluation metrics
Grammatical evolution
Symbolic regression

ABSTRACT

Diabetes mellitus is a metabolic disease involving high blood glucose levels that can lead to serious medical consequences. Hence, for diabetic patients the prediction of future glucose levels is essential in the management of the disease. Most of the forecasting approaches in the literature evaluate the effectiveness of glucose predictors only with numerical metrics. These approaches are limited because they evenly treat all the errors without considering their different clinical impact that could involve lethal effects in dangerous situations such as hypo- or hyperglycemia.

To overcome such a limitation, this paper aims to devise models for reducing high-risk glucose forecasting errors for Type 1 diabetic patients. For this purpose, we exploit a Grammatical Evolution algorithm to induce personalized and interpretable forecasting glucose models assessed with a novel, composite metric to satisfy both clinical and numerical requirements of the estimated predictions.

To assess the effectiveness of the proposed approach, a real-world data set widely used in literature, consisting of data from several patients suffering from Type 1 diabetes, has been adopted. The experimental findings show that the induced models are interpretable and capable of assuring predictions with a good tradeoff between medical quality and numerical accuracy and with remarkable performance in reducing high-risk glucose forecasting errors. Furthermore, their performance is better than or comparable to that of other state-of-the-art methods.

1. Introduction

Diabetes is an incurable disease rapidly spreading over a large part of the population around the world [1]. It is linked to a pancreas malfunction that causes high blood glucose levels over a protracted period. The accumulation of blood glucose can cause damage to multiple organs with long-term severe medical complications [2]. Among the diabetes pathogenesis, Type 1 Diabetes Mellitus (T1DM) occurs when the insulin-producing β-cells of the pancreas are damaged and induce an absolute insufficiency of insulin that provokes chronic hyperglycemia. Currently, the most adopted therapy for T1DM treatment is represented by an open-loop device consisting of a continuous subcutaneous insulin pump directly managed by the patient. This device is coupled with non-invasive Continuous Glucose Monitoring (CGM) systems [3] capable of measuring the current glucose in the subcutaneous tissue.

In recent years, approaches based on Artificial Intelligence (AI) and Machine Learning (ML) have been investigated to provide patients with tools able to effectively predict future glucose levels [4–10]. These tools consider forecasting horizons lasting at most two hours and assist patients in managing insulin every day, given that the involved forecasting horizon typically represents the time a patient needs to decide on the insulin dose to assume following a meal or a hyperglycemic event.

Despite such recent progress, given the complex nature of glucose metabolism, an accurate prediction of future glucose levels remains a challenge for diabetes treatment. This prediction requires particular attention from a clinical viewpoint, as mispredictions in the hypoglycemic and hyperglycemic zones can be fatal for diabetic patients. However, although ML approaches can play a crucial role in improving glucose prediction accuracy, it should be evidenced that most of these approaches adopt measures of performance aiming at only evaluating the numerical accuracy while neglecting the clinical aspects of the estimated predictions. Even the most popular and performing techniques

* Corresponding author.
E-mail addresses: adellacioppa@unisa.it (A. Della Cioppa), ivanoe.defalco@icar.cnr.it (I. De Falco), txkoutny@kiv.zcu.cz (T. Koutny), umberto.scafuri@icar.cnr.it (U. Scafuri), ublm@kiv.zcu.cz (M. Ubl), ernesto.tarantino@icar.cnr.it (E. Tarantino).

https://doi.org/10.1016/j.asoc.2023.110012
1568-4946/

exploited so far in this field, i.e., Deep Neural Networks (DNNs), while exhibiting an appealing numerical accuracy, make use of opaque models that yield difficult explaining what aspects of the model input drive the decisions (explainability) and, more importantly, preclude domain experts from any possibility of understanding how a prediction has been made (interpretability) [11–13]. More specifically, intrinsic interpretability consists of self-explanatory models that feature extremely clear explanations but usually do not have outstanding performance. In contrast, *post-hoc* explainable systems, like DNNs, although keeping high performance, must build a second model that explains their decisions, but these explanations may be poor. When specific reference is made to the medical domain, it is well known in the literature, e.g., [11,12], that black box models are complicated to troubleshoot, which is a remarkably grave problem in this specific field. Moreover, black box models transform computer-aided resolutions into automated resolutions because physicians are not provided with real knowledge of the reasoning processes that take place in black box models. This lack of an explanation on how a prediction is made for a specific patient can limit the use of deep learning models in medical decision support [14]. Attempting to provide *a posteriori* an explanation to black boxes can make this problem even worse because such an attempt might yield deceiving or untrue descriptions, as this process might yield *post-hoc* models that are, at least partially, not the same as the model build internally. These problems are overcome with the use of interpretable machine learning. In the following, the term interpretability is used in the sense described above.

To evaluate the clinical accuracy of a glucose measurement device, analytical accuracy relying on numerical metrics, e.g., Root Mean Square Error (RMSE) or Mean Percentage Absolute Error (MAPE), is not suitable because it often fails to reveal significant outliers. This problem is of paramount importance in hypoglycemia and hyperglycemia situations. Rather, treatment decisions should be considered, which is accomplished through the use of error grid analysis (EGA). This latter consists in identifying a set of risk zones, considering pairs of corresponding estimated - reference values, and computing the percentage of pairs falling in the most dangerous zones. Several EGAs exist in the literature; we can recall here the Clarke Error Grid (CEG) [15], the Parkes Error Grid (PEG) [16], and the Surveillance Error Grid (SEG) [17]. Although useful, these grids are overly simplistic because they only reckon the single glucose values but do not take glucose dynamics into account, meaning with this the way successive values are tied in time. A step forward towards better clinical accuracy assessment consists in the Continuous Glucose-EGA (CG-EGA) [18], where not only errors in rate are considered, but also those in the direction of blood glucose variation.

The situation changes completely when the accuracy should be evaluated of a device that does not aim at measuring current values but rather at performing predictions for future glucose values. In this latter case, all the above-mentioned methods are not appropriate. To evaluate the accuracy in predicting glucose values, the direct use of CEG, PEG, SEG, or CG-EGA methodologies does not allow for correctly assessing the performance of predictors. CEG, PEG, and SEG would suffer from the identical drawback described above, i.e., they do not consider the glucose dynamics. Even the use of the most sensitive measurement estimator, i.e., CG-EGA, for prediction has the critical drawback that the estimation of the rates for changes in glucose values is performed backward, i.e., from past glucose readings, while a predictor should yield glucose estimates ahead in time. This drawback may involve inaccurate predictions of glucose variations that can confuse diabetic patients in understanding the future glycemia evolution and lead to inadequate treatments in risky situations.

To overcome these limitations, in [19] Prediction-Error Grid Analysis (PRED-EGA) was advanced, in which the estimation of change rates takes place on predicted values so that it constitutes a rigorous metric to assess prediction accuracy. It can be seen as a particularization of CG-EGA to the prediction problem.

Nevertheless, up to now, PRED-EGA has only been utilized either in conjunction with other ML methodologies such as regression and neural-network-based models [20], DNNs [21] only to measure *a posteriori* the clinical acceptability of the attained predictions, or directly utilized to enhance the clinical acceptability into DNNs, i.e., Long Short-Term Memory (LSTM), models [22].

In our opinion, a rigorous metric to assess the clinical acceptability in harmful situations, like PRED-EGA, should be exploited to drive the learning process, but a performance measure based on its exclusive use may lead to significant absolute prediction errors.

Starting from the above considerations, the purpose of this paper is:

- to improve the clinical acceptability, also termed clinical quality, of the estimated predictions and, at the same time, taking into account numerical criteria, to downsize the variation range of the errors;
- to automatically extract personalized and interpretable regression models able to estimate glucose levels in future prediction horizons.

The first goal has been accomplished by considering a linear combination of clinical and numerical metrics as an objective function for rating the accuracy of the glucose predictor. This combination allows surmounting the limit of forecasting models assessed by objective functions only based on numerical criteria that assign the same clinical impact to all the prediction errors, which may involve wrong diabetes treatments. The clinical criterion for modeling the glucose predictors is based on PRED-EGA, thus aiming at yielding discovered models more sensitive to risky situations like hypo- and hyperglycemia.

With reference to this first goal, the novelty lies in the fact that, for the first time in the literature, the clinical quality of glucose prediction assessed through PRED-EGA methodology is improved through the use of an evolutionary algorithm. It should be remarked here that, up to now, evolutionary algorithms have been used to explicitly optimize the clinical quality evaluated either by adding a function that penalizes error prediction in the most harmful zones [23] or other error grids as CEG [24,25], which should be accounted for to assess measure precision in devices, though.

The second goal has been fulfilled by exploiting the capability of the Grammatical Evolution (GE) [26] to automatically evolve interpretable regression models. The interpretability allows providing a rationale behind the decision-making process that can help specialists in discerning the mechanisms and, consequently, understanding what is happening. l

As regards this second goal, a further novelty of our paper, tied to the specific GE implementation, is that these regression models are based on the knowledge of current and past glucose measurements, and only the current and future estimated insulin and carbohydrate concentrations to be absorbed.

The discovered personalized models are tested and validated over a real-world data set enclosing the clinical data for several subjects suffering from T1DM.

The paper structure is the following. Section 2 exposes a review of the glucose forecasting-related works in the literature. Section 3 illustrates the GE-based approach to tackle the regression problem. The experimental results are presented, commented on, and compared with other approaches in Section 4. Section 5 reports the final remarks and some hints on future work.

2. Glucose forecasting related works

The numerous forecasting models proposed for diabetes can be classified into physiological, data-driven, and hybrid models [27,28].

The physiological models concern glucose-insulin mathematical models for mimicking human physiological behavior through the description of the glucose dynamics in the course of insulin treatment [29–31]. Unfortunately, these models are often inaccurate due to the necessity of having prior knowledge for setting the physiological constants [32].

The data-driven models provide glucose forecasting by learning patterns from data. Latest developments in data-driven techniques, artificial intelligence, and machine learning have stimulated the research interest in glucose level prediction models as is reported in many recent survey papers [8–10].

Data-driven models typically rely on ML techniques. Several classes of ML methodologies have been utilized for glucose prediction. The techniques employed are feed-forward and recurrent neural networks [33,34], jump and deep neural networks [35–37], autoregressive neural networks and neuro-fuzzy networks [38,39], convolutional neural networks [40,41], support vector regression (SVR) [42], random forests and kernel functions [43], Gaussian processes and self-organizing maps [44], neuroevolution approaches [45], and evolutionary techniques [46–49].

The hybrid models merge physiological models with other techniques for inferring glucose dynamic models. The majority of these hybrid models combine data-driven and compartmental models [23,47,50–54]. Our approach belongs to this latter area.

The ML techniques employed to carry out the prediction evidence that there is no common view about the number and type of diabetes parameters employed as input for training the models. According to the literature reviewed in [8], the set including input parameters such as glucose values, injected insulin, and carbohydrate intake is the most used group of parameters to derive prediction models.

The present paper falls in this set as regards the selection of input parameters and uses a GE as the ML approach to extract the prediction models as in [23–25,47,55–61].

To enhance the performance in most of the above-referenced papers, the standard GE is combined with other approaches or investigated in several adaptations for dealing with different operating conditions. Moreover, the authors consider only numerical metrics as objective functions with some exceptions in monobjective [46,56,57] and multiobjective approaches [24,25] in which a CEG-based metric is included in the objective function.

Differently from the above GE approaches, the clinical accuracy is evaluated in this paper through PRED-EGA able to capture not only the accuracy of the point predictions but also the dynamics of the predicted rate of changes between two consecutive predictions.

We consider an objective function structured as a linear combination to account for the clinical accuracy, assessed in terms of PRED-EGA, while at the same time safeguarding the numerical accuracy that allows reducing the prediction deviations from target values. In this way, our approach can provide a single and effective model to deal with the numerical and clinical perspectives rather than choosing among several equivalent models proposed by Pareto multiobjective approaches.

As far as we know, only few recent papers consider a clinical metric based on PRED-EGA, in addition to the numerical accuracy, to directly evaluate [20,22] or 'a posteriori' estimate the glucose predictors [21].

Without any pretense of being exhaustive, to immediately evidence the positioning of our paper, Table 1 outlines a comparison concerning the main features only versus the above-referenced glucose forecasting techniques using either GE or the same clinical metric. In particular, the comparison is effected in terms of the multiobjective approach adopted, the employment of an evolutionary algorithm to perform the optimization, the explicit use of a clinical metric in the objective function (Cl-Met) and not as 'a posteriori' evaluation tool, and the use of PRED-EGA clinical criterion in such a function. It should be noted that, P indicates methods dealing with a Pareto-front multiobjective approach, while LC refers to a multiobjective approach reduced to a single objective function through an aggregation as a linear combination with positive coefficients.

As it can be observed from Table 1, the optimization of PRED-EGA through an evolutionary algorithm is exclusive to our paper.

The last three papers in Table 1, namely [20–22], are interesting to test the effectiveness of our proposal.

De Bois et al. in [20] presented the numerical accuracy and the clinical acceptability of nine models taken from literature. The approach, named GLYFE (GLYcemia Forecasting Evaluations), includes as models a baseline comparison (Base), a polynomial regression model (Poly), two Auto-Regressive models (AR, ARX), two more complex non-linear regression models, i.e., Support Vector Regression (SVR) and Gaussian Process (GP), and three models relying on neural networks, namely Extreme Learning Machine network (ELM), Feed-forward Neural network (FFN), and Long Short-term Memory (LSTM).

In the other paper [22], the same authors tried to enhance the clinical quality of the deep models by enriching LSTM with new loss functions: the coherent mean squared error (cMSE) that is the Mean Squared Error (MSE) weighted by the MSE of the predicted variations, and the coherent mean squared glycemic error (gcMSE) based on an error grid analysis to deal with the prediction errors and their variations during the training. In particular, this last function relies on weighting coefficients of a metric able to improve the clinical quality of the model to the detriment of numerical accuracy. Therefore, the authors proposed the Progressive Improvement of the Clinical Acceptability (PICA) algorithm by considering these two contrasting objectives. Further details can be found in [22].

Dudukcu et al. [21] exploited a combination of an LSTM, Gated Recurrent Units, and WaveNet DNNs for glucose prediction by only using the glucose values of patients' history for the prediction.

These three papers will represent the state-of-the-art comparison carried out in Section 4.3.

3. Grammatical evolution for glucose forecasting

The proposed evolutionary approach falls within data-driven models that exploit time series coming from CGM systems to forecast future glycemic trends for T1DM patients. The problem can be seen as a multivariate time series regression. Namely, let us suppose that we have measured the glucose level $G(t)$, the administered insulin $U(t)$, and the ingested amounts of carbohydrates $D_g(t)$.

It is to note that the injected insulin corresponding to insulin boluses plus insulin basal and the assumed carbohydrates are discrete signals that need to be converted into continuous signals to estimate their effects on the glycemic trend over time. The preprocessing of the injected insulin boluses is performed based on the Hovorka model [62] delineating the absorption rate of the injected insulin through a two-compartment chain. This model allows adding the signal representing the absorption rate of the boluses to the signal describing the absorption rate of subcutaneously administered long-acting insulin. Specifically, the model for insulin absorption is:

$$\frac{dS1}{dt} = U(t) - \frac{S1}{t_{\max_I}} \tag{1}$$

Table 1
Comparison of the main features of the referenced forecasting techniques in terms of the multiobjective approach.

Paper	Multiobjective	EA	Cl-Met	PRED-EGA
Hidalgo et al. [55]	No	Yes	No	No
Contreras et al. [23]	No	Yes	Yes	No
Velasco et al. [46]	No	Yes	Yes	No
Hidalgo et al. [47]	No	Yes	No	No
Velasco et al. [56]	No	Yes	Yes	No
Velasco et al. [57]	No	Yes	Yes	No
De Falco et al. [48]	No	Yes	No	No
De Falco et al. [49]	No	Yes	No	No
Lourenco et al. [58]	No	Yes	No	No
Hidalgo et al. [59]	No	Yes	No	No
Joedicke et al. [24]	Yes (P)	Yes	Yes	No
De Falco et al. [60]	No	Yes	No	No
Contador et al. [25]	Yes (P)	Yes	Yes	No
De Falco et al. [61]	No	Yes	Yes	No
De Falco et al. [45]	No	Yes	No	No
De Bois et al. [20]	No	No	Yes	Yes
Dudukcu et al. [21]	No	No	No	No
De Bois et al. [22]	No	No	Yes	Yes
This paper	Yes (LC)	Yes	Yes	Yes

$$\frac{dS2}{dt} = \frac{S1 - S2}{t_{\max_I}} \tag{2}$$

in which $S1$ and $S2$ are the two-compartment chain for modeling the absorption of subcutaneously infused short-acting insulin, $U(t)$ [mU min^{-1}] is the amount of administered insulin, $t_{\max_I} = 55$ [min] is the constant indicating the time-to-maximum insulin absorption, $S1(t)$ [mU] and $S2(t)$ [mU] are the amounts of insulin in the two compartments. Then the plasma insulin concentration I [mU l^{-1}] is described as:

$$\frac{dI}{dt} = \frac{S2}{V_I \cdot t_{\max_I}} - k_e \cdot I \tag{3}$$

where $k_e = 0.138$ [min^{-1}] is the fractional elimination rate of the insulin from plasma and $V_I = 0.12$ [l kg^{-1}] is the insulin distribution volume. All the considered constants are taken from Hovorka's model [63].

Concerning the carbohydrate intake, in the presence of a meal, the gut absorption rate is modeled according to [62] as:

$$C(t) = \frac{D_g \cdot A_g \cdot t \cdot e^{-t/t_{\max}}}{t_{\max}^2} \tag{4}$$

where $t_{\max} = 40$ [min] is the time-of-maximum appearance rate of glucose in the accessible compartment, D_g is the amount of digested carbohydrates, and $A_g = 0.8$ is the carbohydrates bioavailability [64]. This function quickly augments after the meal and then decreases to 0 in 2–3 h. Outside such a time span, the missing carbohydrate values are filled with zeros.

At the end of the preprocessing, by integrating Eq. (3) and exploiting Eq. (4), we have two signals, discretized every Δt minutes, for the absorbed insulin and carbohydrates, i.e., $I(t)$ and $C(t)$, respectively. More specifically, when there is an insulin delivery or carbohydrate intake event at time t, their absorbed amounts are spread over time through Eqs. (3) and (4) from the current time t ahead and, eventually, added to the residual quantity computed by the compartment model in the past.

As a consequence, by considering the values of $G(t)$ every Δt minutes in a time offset of $k\Delta t$ minutes before the current instant t, and the values of $I(t)$ and $C(t)$ every Δt minutes in a time offset of $h\Delta t$ minutes after the current instant t, we search for an explicit regression model able to forecast the future glucose value $\widehat{G}(t + h\Delta t)$ at a prediction horizon $h\Delta t$:

$$\begin{aligned}\widehat{G}(t + h\Delta t) = {} & \Gamma\big(G(t), G(t - \Delta t), \ldots, G(t - k\Delta t)\big) \\ & - \Theta\big(I(t), I(t + \Delta t), \ldots, I(t + h\Delta t)\big) \\ & + \Omega\big(C(t), C(t + \Delta t), \ldots, C(t + h\Delta t)\big)\end{aligned} \tag{5}$$

where Γ, Θ and Ω are expressions on G, I, and C, respectively.

Note that, analogously to the work by Hidalgo et al. [59], the model (Eq. (5)) we look for exploits past values for glucose while, differently from it, only the amounts of insulin and carbohydrates yet to be absorbed, and evaluated through the compartmental model illustrated above, are taken into account. This choice comes from the consideration that past amounts of insulin and carbohydrates have been absorbed in the current time t; therefore, the current glucose value includes the effects of their absorption.

To extract an explicit regression model from the available data, we exploit an evolutionary approach based on a GE algorithm [26]. Such an algorithm evolves a population of regression models, i.e., expressions, obtainable through a context-free Backus Naur Form-style (BNF) grammar to optimize (minimize in this case) an objective function.

3.1. Objective function

The measurement of model prediction accuracy is indispensable for assessing the results' reliability and the medical outcome of diabetes therapy. Several metrics exist for the performance evaluation of the forecasting models [8]. These metrics can be subdivided into mathematical and clinical evaluation criteria, accounting for numerical and clinical accuracy.

The mathematical criteria are employed to assess the numerical accuracy without providing insights related to the clinical importance. Within the category of mathematical criteria, we have considered the RMSE and the MAPE.

The most employed evaluation criterion to estimate the clinical accuracy of the predictions is the error grid analysis [8]. Within this work, the clinical accuracy of the regression models has been assessed through PRED-EGA [65,66]. PRED-EGA combines two graphical grid representations: Point-Error Grid Analysis (P-EGA) for the clinical accuracy of point predictions and Rate-Error Grid Analysis (R-EGA) for that of the predicted rate of changes between two consecutive predictions. As a function of the couple, G and $\widehat{G}$, each grid is partitioned into five zones from A (best) to E (worst) that represent an increasing risk category when making predictions. Specifically, zone A includes accurate estimations and zone B refers to inaccurate predictions with little outcome on diabetes management. Zone C entails potentially risky treatments, zone D serious unpredicted hypo- or hyperglycemia events, while prediction in zone E corresponds to unpredicted high-risk situations. According to the real glycemia regions (hypoglycemia (G $<$ 70 [mg dl^{-1}]), euglycemia (70 [mg dl^{-1}] $\leq$

Table 2
Protected functions used in the grammar.

Function	Protected function
plog(x)	$\log(1+\lvert x\rvert)$
psqrt(x)	$\sqrt{\lvert x\rvert}$
aq(x, y)	$\frac{x}{\sqrt{1+y^2}}$

$G \leq 180$ [mg dl^{-1}]), and hyperglycemia ($G > 180$ [mg dl^{-1}])), both markers of P-EGA and R-EGA are employed to establish whether a prediction is an Accurate Prediction (AP), a Benign Error (BE), or an Erroneous Prediction (EP). A model with acceptable clinical quality should have a high AP rate and a low EP rate in the dangerous hypo and hyper zones.

To assess the forecasting model taking into account the clinical quality without neglecting the numerical accuracy, the following linear combination of objective functions to be minimized has been devised:

$$\Phi_3(G, \widehat{G}) = \alpha \cdot \Phi_1(G, \widehat{G}) + \Phi_2(G, \widehat{G}) \tag{6}$$

where

$$\Phi_1(G, \widehat{G}) = \mathrm{RMSE}(G, \widehat{G}) \tag{7}$$

only relies on a numerical evaluation, and

$$\Phi_2(G, \widehat{G}) = \mathrm{EP}_i + \mathrm{EP}_e + \mathrm{EP}_h \tag{8}$$

only focuses on clinical evaluation criteria concerning EP_i, EP_e, and EP_h representing the ratios between the number of erroneous predictions and the number of samples in hypoglycemia, euglycemia, and hyperglycemia zones, respectively. This last assessment function assumes noticeable importance in hypoglycemia and hyperglycemia regions since the risk associated with erroneous predictions in such regions can be critical. This reduces the eventuality that diabetic patients are not warned of these potentially dangerous events.

As regards Φ_3, being the RMSE numerical metric value much greater than $\mathrm{EP}_i + \mathrm{EP}_e + \mathrm{EP}_h$, to avoid the numerical component can be predominant, RMSE is multiplied by a modulation factor $0 < \alpha < 1$. The clinical component aims to drive the objective function in providing acceptable predictions that make sense from a medical point of view; this can be obtained by avoiding mispredictions before all in the most critical hypoglycemia and hyperglycemia regions. The numerical component, in its turn, allows retaining the capability to lower the absolute prediction errors in all the zones.

Henceforth the objective functions Φ_1 and Φ_2 are employed in this paper only for comparison purposes.

3.2. Grammar

The syntax of the expressions to be evolved by the GE is specified by the context-free grammar depicted in Fig. 1, where $\langle gluc \rangle$ represents the glucose levels in the past while $\langle ins \rangle$ and $\langle cho \rangle$ are those of insulin and carbohydrates in the future, respectively. In our grammar, psqrt and plog are protected functions that return the square root of the absolute value of the argument and the logarithm of the summation of 1 and the absolute value of the argument, respectively, while aq represents the protected analytic quotient operator [67]. Table 2 shows the protected functions used in the grammar.

4. Experimental results

4.1. Experimental framework setting

The experiments are conducted on the Ohio T1DM data set, released in 2018 [68], that gathers eight weeks of data for six T1DM patients subject to insulin pump therapy. This specific data set has been chosen to perform a comparison with other papers in the literature that employ the same values for past offset and predicting horizon, and to evaluate the effectiveness of the discovered forecasting models with the same numerical and clinical metrics. The data set includes glucose measurements obtained by a CGM system with a sampling interval of $\Delta t = 5$ minutes and self-measurements by finger pricking; bolus and basal insulin doses; self-reported information concerning mealtimes and carbohydrate estimates, times of sleep, physical activity, work, illness; and several other physical variables, including aggregations of heart rate, galvanic skin response, skin temperature, air temperature, and psychological stress.

From all the variables in the data set, we only examined those included in a standard CGM profile, i.e., the glucose level, the injected insulin (basal plus boluses), and the carbohydrates assumed during the day. Since the Ohio T1DM data are not specific to nights, instead they span over all 24 h, and since the glucose level is affected by physical and psychological variations [69], these data are adequate to evaluate the robustness of the model under different disturbances.

To perform supervised learning, the data series of each patient is partitioned into training and testing sets used to extract the model during the learning phase and assess its quality over unseen samples. The number of training and testing items for each patient is reported in [68].

During the preprocessing, the problem of missing glucose data arises. We have decided to throw away samples with missing glucose readings in training and testing sets to prevent the forecasting models from being the result of artificial observations.

4.2. Findings

PonyGE2, a freely downloadable and patent-free GE implementation in Python, has been used for the regression problem [70]. The parameters used for all the experimental trials have been set after a preliminary tuning: population size and generations equal to 200 and 1,000, respectively; codon size equal to 100,000, tournament selection with size 4, mutation probability equal to 10%, one-point crossover probability equal to 90%, int flip per codon mutation with one mutation event, and Position Independent Grow method for the individual initialization.

Each of the three fitness functions defined in Section 3.1 has been used as the objective function. This leads to the design of three GE algorithms: the generic algorithm containing the fitness function Φ_i is denoted as GE_{Φ_i}. The modulation factor α related to GE_{Φ_3} is set to 0.025 [dl mg^{-1}] after performing a preliminary tuning.

The adopted prediction horizon is $h\Delta(t) = 30$ minutes because the prediction accuracy becomes worse and less reliable as the forecasting horizon increases [47]. A horizon longer than 30 min, e.g., two or four hours, is only practical for time spans that contain almost steady-state situations as nocturnal predictions when sleeping. Any external event can cause a significant and unpredictable glucose level change during these long intervals. The considered past time offset is $k\Delta(t) = 30$ minutes for the historical samples exploited for the prediction. The time interval for the historical data is chosen based on the consideration that 30-minute data in the past are enough to make an effective prediction [60].

For each patient, indicated with the identifier ID, twenty runs have been carried out to reduce the randomness in the initialization of the GE algorithm. The evaluation is performed in all instances for which a glucose measurement is available over the prediction period. The average outcomes for each run and patient are evaluated at the end of the evolution.

$\langle glucose \rangle ::= (\langle e\_gluc \rangle) + \langle d \rangle.\langle d \rangle * \text{abs}(\langle e\_cho \rangle) - \langle d \rangle.\langle d \rangle * \text{abs}(\langle e\_ins \rangle)$

$\langle e\_gluc \rangle ::= (\langle e\_gluc \rangle \langle op \rangle \langle e\_gluc \rangle) \mid \text{aq}(\langle e\_gluc \rangle, \langle e\_gluc \rangle) \mid \langle func \rangle(\langle e\_gluc \rangle) \mid \langle gluc \rangle \mid \langle number \rangle$

$\langle e\_ins \rangle ::= (\langle e\_ins \rangle \langle op \rangle \langle e\_ins \rangle) \mid \text{aq}(\langle e\_ins \rangle, \langle e\_ins \rangle) \mid \langle func \rangle(\langle e\_ins \rangle) \mid \langle ins \rangle \mid \langle number \rangle$

$\langle e\_cho \rangle ::= (\langle e\_cho \rangle \langle op \rangle \langle e\_cho \rangle) \mid \text{aq}(\langle e\_cho \rangle, \langle e\_cho \rangle) \mid \langle func \rangle(\langle e\_cho \rangle) \mid \langle cho \rangle \mid \langle number \rangle$

$\langle op \rangle ::= + \mid - \mid *$

$\langle func \rangle ::= \text{plog} \mid \text{psqrt} \mid \sin \mid \tanh \mid \exp$

$\langle gluc \rangle ::= \text{G(t)} \mid \text{G(t-}\Delta t\text{)} \mid \ldots \mid \text{G(t-}k\Delta t\text{)}$

$\langle ins \rangle ::= \text{I(t)} \mid \text{I(t+}\Delta t\text{)} \mid \ldots \mid \text{I(t+}h\Delta t\text{)}$

$\langle cho \rangle ::= \text{C(t)} \mid \text{C(t+}\Delta t\text{)} \mid \ldots \mid \text{C(t+}h\Delta t\text{)}$

$\langle number \rangle ::= \langle d \rangle.\langle d \rangle \mid -\langle d \rangle.\langle d \rangle$

$\langle d \rangle ::= [0, \ldots, 99]$

Fig. 1. The grammar for the glucose forecasting model (Eq. (5)).

4.2.1. Algorithm performance

The quality of the forecasting models is estimated in terms of both clinical and numerical accuracy. To be coherent with most of the results in the literature, the clinical accuracy is expressed in percentage, while the numerical accuracy is reported in [mg dl^{-1}] for RMSE and percentage for MAPE. Namely, for each of the three GE algorithms, Table 3 reports for each subject the average percentages on the testing set of the nine PRED-EGA clinical parameters together with their standard deviations over the 20 runs. A further row shows the values of the parameters averaged over the six subjects. High AP, low BE, and EP rates attest to acceptable clinical quality.

As concerns the numerical accuracy, i.e., RMSE and MAPE, for each GE version, Table 4 reports the results on the testing set for each subject in terms of the average values with the related standard deviations over the 20 runs. The last row shows the numerical values averaged over the six subjects.

About the quality of the GE-based forecasting model for the numerical (only evaluated according to RMSE) and the complementary clinical accuracy, from Table 3, GE_{Φ_1} turns out to have the worst behavior from a clinical viewpoint witnessed by a low value of the average AP rate (61.24%) and a very high value of the average EP rate (35.28%) in the critical hypoglycemia region. In contrast, GE_{Φ_1} presents the best average numerical accuracy over all the patients (21.41 [mg dl^{-1}]) as shown in Table 4.

A dual behavior can be observed as concerns the clinical quality of GE_{Φ_2} that exhibits the best average AP percentage and very low average EP rate in the hypoglycemia region. This algorithm is characterized by a noticeable worsening in the average numerical accuracy (41.40 [mg dl^{-1}]) as evidenced in Table 4. It is worth noting that the good performance in the critical hypoglycemia region of this algorithm is of paramount importance from a clinical perspective. On all the twenty runs, it never makes serious errors (EP = 0.0%) over three subjects.

The last algorithm GE_{Φ_3} combines the numerical and clinical qualities. It presents acceptable values for the average AP and BE, but before all low EP rates in all the regions and a good value for the average numerical accuracy (24.37 [mg dl^{-1}]). This demonstrates that the predictor assessed through a combination of numerical and PRED-EGA-based clinical metrics still allows for reducing the predicted errors in high-risk situations while safeguarding the numerical accuracy of the predictions. It could represent a good compromise when we want to privilege the medical aspect without penalizing the numerical one.

About the performance stability[1] of the algorithms, it can be inferred by looking at the low standard deviations over the different patients for each region.

The observations above evidence very good features of our forecasting tools in different operating conditions.

It would be too lengthy to describe and analyze the models obtained for all the investigated subjects. Therefore, in the rest of this paper, we will only consider one of them as an exemplary case. To avoid discussing the behavior of borderline subjects, we have chosen subject 559 because they exhibit an intermediate number of hypoglycemic, euglycemic, and hyperglycemic cases among all the patients.

In Fig. 2 we report the PRED-EGA (P-EGA and R-EGA) on the testing set regarding the patient with ID=559 for the algorithms GE_{Φ_1} (top), GE_{Φ_2} (middle), and GE_{Φ_3} (bottom). The R-EGA panes show that the algorithm GE_{Φ_2} makes a lower number of serious prediction errors (represented by red points), especially in the dangerous zone E where predicted rates of change are opposite to what happens. This is to be expected, as these algorithms are designed to avoid predictions in E. Instead, algorithms GE_{Φ_1}

[1] With the term stability we refer to the characteristic of an evolutionary algorithm regarding its sensitivity to perturbations of the initial conditions, i.e., the randomness related to the initialization of the initial population.

Table 3
Average clinical quality of glucose predictive models (with standard deviation) on the testing set for the three algorithms.

	ID	PRED-EGA								
		Hypoglycemia			Euglycemia			Hyperglycemia		
		AP (%)	BE (%)	EP (%)	AP (%)	BE (%)	EP (%)	AP (%)	BE (%)	EP (%)
GE_{Φ_1}	559	85.48 (0.69)	1.39 (0.00)	13.13 (0.69)	94.74 (0.18)	3.00 (0.19)	2.25 (0.10)	87.65 (0.19)	8.83 (0.24)	3.53 (0.13)
	563	49.64 (1.56)	0.36 (1.56)	50.00 (2.26)	91.69 (0.13)	6.13 (0.13)	2.19 (0.08)	88.44 (0.21)	8.37 (0.22)	3.20 (0.13)
	570	16.67 (11.91)	10.00 (4.84)	73.34 (8.89)	96.07 (0.60)	3.67 (0.62)	0.25 (0.13)	92.01 (0.44)	6.63 (0.29)	1.36 (0.18)
	575	64.89 (2.28)	3.88 (0.83)	31.23 (2.28)	91.15 (0.15)	5.32 (0.17)	3.54 (0.16)	85.56 (0.41)	7.44 (0.36)	7.00 (0.21)
	588	90.00 (23.80)	0.00 (0.00)	10.00 (23.80)	89.71 (0.24)	7.26 (0.25)	3.04 (0.21)	88.23 (0.30)	9.35 (0.26)	2.42 (0.17)
	591	60.73 (1.91)	5.26 (0.00)	34.01 (1.91)	85.81 (0.16)	9.81 (0.20)	4.37 (0.15)	82.47 (0.25)	13.98 (0.20)	3.56 (0.22)
	Avg	61.24	3.48	35.28	91.53	5.87	2.61	87.39	9.10	3.51
GE_{Φ_2}	559	94.93 (1.26)	2.78 (0.00)	2.29 (1.26)	94.57 (0.16)	3.08 (0.16)	2.35 (0.06)	87.17 (0.31)	8.89 (0.33)	3.94 (0.13)
	563	92.86 (0.00)	7.14 (0.00)	0.00 (0.00)	91.61 (0.12)	6.03 (0.30)	2.37 (0.25)	87.26 (0.38)	7.84 (0.35)	4.90 (0.32)
	570	87.78 (5.98)	12.22 (5.98)	0.00 (0.00)	95.67 (0.37)	3.87 (0.28)	0.46 (0.20)	91.20 (0.34)	6.87 (0.30)	1.93 (0.10)
	575	89.89 (1.30)	6.31 (0.76)	3.81 (1.60)	90.98 (0.22)	5.51 (0.17)	3.52 (0.15)	83.84 (0.70)	7.91 (0.50)	8.24 (0.46)
	588	100.00 (0.00)	0.00 (0.00)	0.00 (0.00)	89.17 (0.23)	7.34 (0.17)	3.49 (0.21)	85.68 (0.42)	9.30 (0.26)	5.02 (0.30)
	591	87.69 (2.00)	6.46 (1.07)	5.85 (1.23)	85.39 (0.28)	10.20 (0.24)	4.40 (0.37)	79.36 (3.40)	13.92 (2.46)	6.72 (0.98)
	Avg	92.19	5.82	1.99	91.23	6.01	2.76	85.75	9.12	5.12
GE_{Φ_3}	559	90.28 (0.00)	1.39 (0.00)	8.33 (0.00)	94.87 (0.07)	2.93 (0.08)	2.21 (0.04)	87.64 (0.12)	8.79 (0.12)	3.56 (0.11)
	563	83.93 (7.45)	7.14 (0.00)	8.93 (7.45)	91.71 (0.16)	6.09 (0.15)	2.20 (0.15)	88.21 (0.26)	8.33 (0.32)	3.46 (0.31)
	570	86.11 (6.92)	11.11 (0.00)	2.78 (6.92)	95.61 (0.39)	3.77 (0.35)	0.62 (0.14)	91.58 (0.40)	6.95 (0.36)	1.47 (0.11)
	575	75.75 (1.82)	4.96 (0.49)	19.29 (1.76)	91.03 (0.24)	5.43 (0.20)	3.54 (0.13)	85.18 (0.28)	7.59 (0.27)	7.24 (0.23)
	588	100.00 (0.00)	0.00 (0.00)	0.00 (0.00)	89.78 (0.31)	7.11 (0.18)	3.11 (0.26)	88.09 (0.41)	9.34 (0.33)	2.57 (0.19)
	591	79.91 (1.62)	5.50 (0.43)	14.59 (1.45)	85.55 (0.18)	9.89 (0.14)	4.56 (0.15)	82.16 (0.31)	13.76 (0.31)	4.08 (0.26)
	Avg	86.00	5.02	8.99	91.43	5.87	2.71	87.14	9.13	3.73

Table 4
Average numerical accuracy (with standard deviation) on the testing set for the three algorithms.

ID	GE_{Φ_1}		GE_{Φ_2}		GE_{Φ_3}	
	RMSE ([mg dl^{-1}])	MAPE (%)	RMSE ([mg dl^{-1}])	MAPE (%)	RMSE ([mg dl^{-1}])	MAPE (%)
559	21.98 (0.11)	10.24 (0.09)	41.00 (1.64)	25.76 (1.28)	25.50 (0.31)	13.22 (0.28)
563	19.93 (0.18)	8.95 (0.06)	38.64 (2.15)	21.66 (1.48)	24.34 (1.38)	11.40 (0.67)
570	18.32 (0.17)	6.69 (0.13)	43.92 (2.36)	20.13 (1.33)	20.64 (0.59)	7.37 (0.31)
575	24.69 (0.51)	11.54 (0.24)	36.43 (3.40)	21.55 (2.89)	25.11 (0.52)	11.75 (0.57)
588	19.94 (0.22)	8.78 (0.14)	46.24 (1.87)	24.38 (1.12)	24.61 (1.28)	11.10 (0.70)
591	23.60 (0.12)	13.22 (0.11)	42.14 (4.06)	27.52 (2.48)	26.00 (0.31)	14.28 (0.20)
Avg	21.41	9.90	41.40	23.50	24.37	11.52

and GE_{Φ_3} have a greater amount of non-dramatic errors (orange points) in the C zone of the R-EGA, so they show over-correction for changes, which could lead to over-treatment problems. This problem is less grave for GE_{Φ_2}, for which a higher number of points lies in the D zone, meaning that it is less effective at detecting rapid glucose falls or, more abundantly, rises, which may cause problems as well.

As anticipated in Section 3.1, in Fig. 3 we can appreciate the advantage of using a combination of clinical and numerical criteria in the predictor assessment. The predictor based only on GE_{Φ_1} is capable of yielding the estimated predictions enough adherent to the actual values. Nevertheless, evaluating the prediction errors independently of the risk zone, this predictor is inappropriate from a clinical perspective because it assigns the same clinical risk to all the prediction errors. Therefore, the resulting forecasting model can be unable to timely advise the patient of all the possible adverse hypoglycemic or hyperglycemic events. It must be considered that some prediction errors in harmful situations, like hypo- or hyperglycemia, are potentially very dangerous. The predictor assessed by clinical metrics (GE_{Φ_2}) prevents this flat evaluation. As can be observed from the figure, this model tends to markedly underestimate hypoglycemic values so over-alerting the patient. Simultaneously, it does not overestimate hyperglycemic values so neglecting possible dangerous situations that could necessitate corrective actions. The proposed linear combination of objective functions, i.e., GE_{Φ_3}, inheriting mathematical features, allows for reducing the absolute prediction errors in all the zones. At the same time, the clinical component permits a less marked underestimation in the hypoglycemic zone and a better estimation in the hyperglycemic region to warn the patient in time to take the right decision in both high-risk situations.

4.2.2. Discovered personalized models

According to the Eq. (5), the GE-based evolutionary algorithm allows extracting an intrinsically interpretable personalized model in the form:

$$\widehat{G}(t+30) = \Gamma(G) - \Theta(I) + \Omega(C) \tag{9}$$

In the case of patient 559, the above forecasting model results in the following three models (the unit is mmol/l), i.e., Eqs. (10), (11), and (12), depending on the algorithms used, i.e., GE_{Φ_1}, GE_{Φ_2}, and GE_{Φ_3}:

$$GE_{\Phi_1} \Rightarrow \begin{cases} \Gamma(G) = G(t) \\ \Theta(I) = 0.70 \cdot |I(t)| \\ \Omega(C) = 0.75 \cdot |\exp(\sin(C(t+15)))| \end{cases} \tag{10}$$

$$GE_{\Phi_2} \Rightarrow \begin{cases} \Gamma(G) & = G(t) \\ \Theta(I) & = 53.29 \cdot |\mathrm{aq}(61.36, 18.80 \\ & \quad \cdot (78.77 + \sin(I(t+20))))| \\ \Omega(C) & = 26.3 \cdot |\mathrm{aq}(C(t+30), -29.97)| \end{cases} \tag{11}$$

$$GE_{\Phi_3} \Rightarrow \begin{cases} \Gamma(G) = G(t) \\ \Theta(I) = 0.60 \cdot |I(t)| \\ \Omega(C) = 0.83 \cdot |\mathrm{aq}(C(t), C(t+5) + C(t+15) \\ \qquad - \ \mathrm{pqsrt}(\exp(C(t+30))))| \end{cases} \tag{12}$$

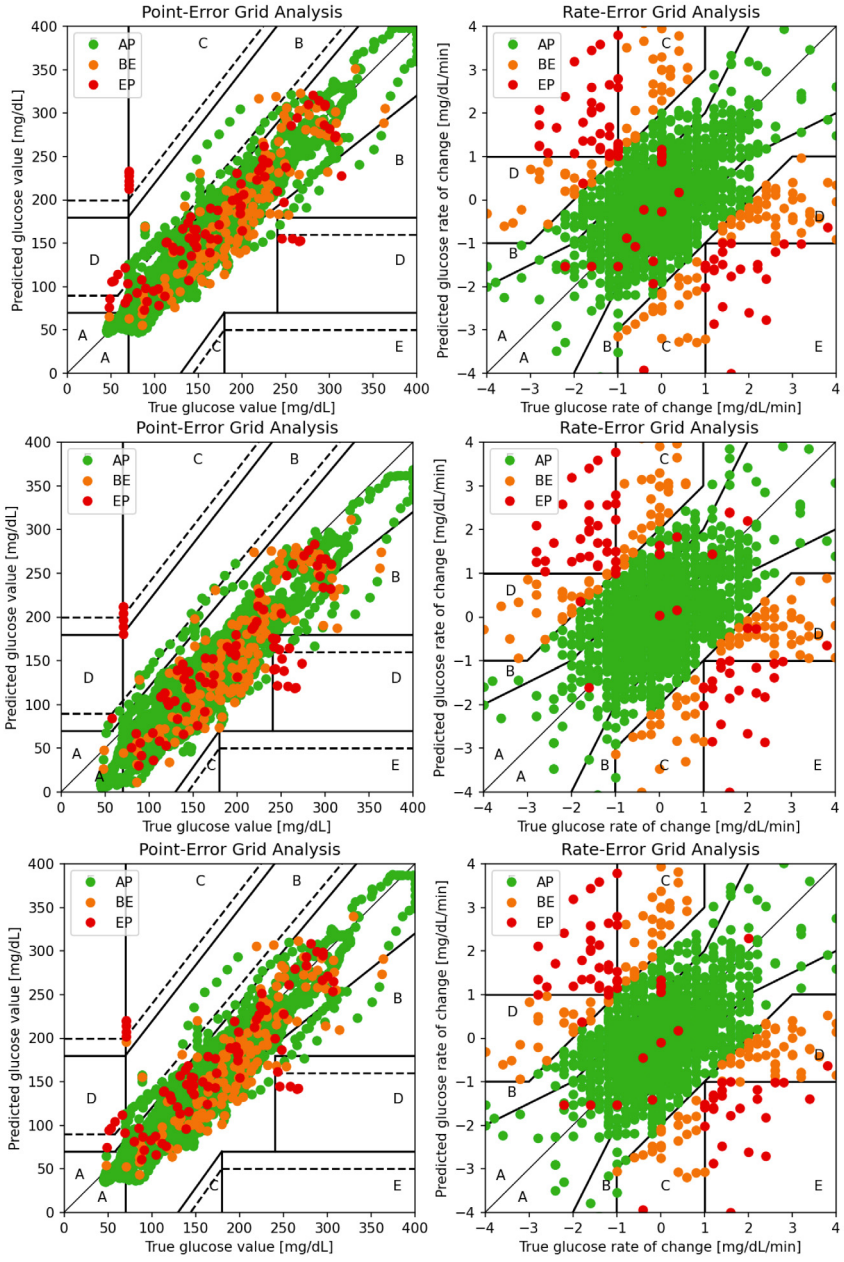

Fig. 2. The PRED-EGA on the testing set for patient 559: GE_{Φ_1} (top), GE_{Φ_2} (middle), and GE_{Φ_3} (bottom).

In general, insulin absorption causes a glucose concentration decrease, while carbohydrate absorption produces an increase in this concentration. It is worth noting that the structures of the models detected for all the other patients are similar. In our opinion, these explicit personalized models could offer an additional aid to physicians in calibrating a personalized treatment for each patient once they are transformed into a user-friendly tool to observe individual reactions to events such as food intake. As it might be expected, all the extracted models evidence an automatic selection of the current glucose amount $G(t)$ with the addition of other amounts of $I(t)$ and $C(t)$ to be absorbed in time steps successive to t.

Fig. 4 shows the insulin and carbohydrate signals and their combined effect after the rearrangement performed by the different algorithms. These graphics allow us to consider the dynamics induced by insulin and carbohydrate absorption on the glycemic trend of the respective discovered models. The rearranged carbohydrate and insulin signals and their combination are similar for GE_{Φ_1} and GE_{Φ_3}. In particular, this latter ranges between positive and negative values resulting in a real adjustment of the glucose concentration. The carbohydrate signal for GE_{Φ_1} is always greater than zero with a minimum at 13.5 mg/dl, while this does not occur for GE_{Φ_2} and GE_{Φ_3}, thus evidencing the limits of a purely numerical approach. As regards GE_{Φ_2}, the whole effect results in a strong downward shift of the glucose signal. Such a behavior depends on the very low number of hypoglycemic events with respect to euglycemic and hyperglycemic events for the subject 559 and on the specific fitness that only takes into account the sum of the error percentages on the three zones. Vice versa, in the case of a dataset with fewer hyperglycemic than hypoglycemic and euglycemic cases, the combined signal would only have positive values. Nevertheless, the prediction of high-risk glucose events could be penalized with a dataset with fewer cases in the euglycemic zone than in the other two zones. The problem cannot be easily solved using a weighted sum since it is difficult to establish how to effectively balance the different risk errors. This

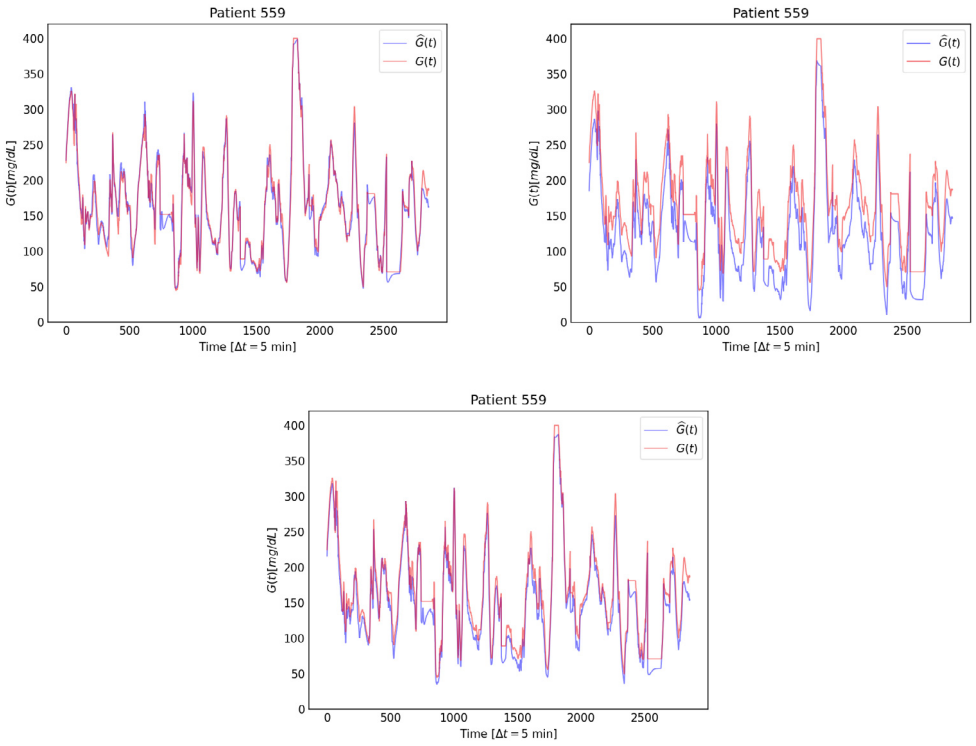

Fig. 3. The glucose forecasting results (in red and blue the actual $G(t)$ and the estimated $\widehat{G}(t)$ respectively) on the testing set for patient 559: GE_{Φ_1} (top left), GE_{Φ_2} (top right), and GE_{Φ_3} (bottom).

behavior of GE_{Φ_2} algorithm evidences very well the drawbacks of a fitness function using just the PRED-EGA. On the contrary, GE_{Φ_3} has none of the limitations highlighted above. Indeed, the carbohydrate signal correctly ranges from zero to positive values, and the combined signal involves an acceptable underestimation in the hypoglycemic zone. Finally, it is to remark that a similar behavior holds for all the patients in the dataset.

4.3. Comparison with literature

To perform a fair comparison, we only consider papers in which the forecasting model is tested over all the six patients included in the 2018 version of the OhioT1DM data set. Many papers exist in the literature facing this data set, yet the vast majority only report results in terms of numerical accuracy, whereas a much lower number contains results in terms of clinical accuracy. To our knowledge, just three papers investigate both numerical and clinical accuracy relying on PRED-EGA, i.e., [20–22]. Hence, in the following Tables 5 and 6, the results on the testing set of our three GE approaches are compared with those presented in De Bois et al. [20], with the best approach outlined in Duducku et al. [21], and the three best deep models included in De Bois et al. [22]. Specifically, these three models are $gpcLSTM_{CA}$ with the lowest EP value in the hypoglycemia region and $gpcLSTM^*_{CA}$ with the lowest EP value in the hyperglycemia region that assess the prediction based on clinical metrics, and $gpcLSTM^*_{PICA}$ using the PICA algorithm which represents a compromise between the numerical and clinical metrics. As concerns Dudukcu et al. [21] that exploited all the 12 patients included in the 2020 version of the updated OhioT1DM data set, we only consider the six patients investigated by De Bois et al. and also in the current paper. Therefore, the best average results for these six patients achieved in correspondence with the most performing approach over the training set, namely LSTM + WaveNet + GRU (LWG) fusion models, are reported in the tables.

For the sake of comparison, the results of our models on the testing are those corresponding to the best objective function value over the training set for each patient. Analogously to the other models, the tables show the averages of these results over all the six patients.

The overall comparison is carried out in terms of clinical criteria, i.e., PRED-EGA (Table 5), and numerical metrics, i.e., RMSE and MAPE (Table 6). The best clinical and numerical metrics values with their corresponding standard deviations are shown in bold.

The compared approaches are grouped regarding the optimization function used to make the predictions, i.e., based only on numerical metrics (G1), only on clinical evaluation criteria (G2), or on the combination of these two metrics (G3).

Looking at Table 5 related to the clinical quality, our algorithm GE_{Φ_1} has a performance that is the best in the hypoglycemia and hyperglycemia regions, and not too far from the best in the euglycemia region with respect to the other models in the same group. In general, by comparing the three algorithms focusing on improving the clinical quality in the G2 group, i.e., $gpcLSTM_{CA}$, $gpcLSTM^*_{CA}$, and our algorithm GE_{Φ_2}, this latter has the best average rate in the hypoglycemia region as regards the AP and, most importantly, the EP which represents a critical issue for evaluating the clinical quality of the predictions. Furthermore, GE_{Φ_2} has similar results in the other two regions. Our combined algorithm GE_{Φ_3} performs much better than the analogous algorithm $gpcLSTM^*_{PICA}$ in the critical hypoglycemia region with comparable results for the euglycemia and hyperglycemia regions.

To reduce the clinical risk, our attention has been concentrated on searching for models capable of lowering the erroneous predictions which involve danger for the patient independently of the occurrence zone. Naturally, errors in the hypoglycemic region could result in wrong treatment decisions with potentially short-term fatal implications. Nevertheless, erroneous predictions in hyper- or euglycemic zones are also risky in the medium or long term if considered reliable estimations. To better highlight the ability of our forecasting models in reducing prediction errors,

Table 5
Average clinical quality (with standard deviation) on the testing set computed by averaging the best glucose predictive models for each patient on the training set.

PRED-EGA												
Group	Papers	Models	Hypoglycemia			Euglycemia			Hyperglycemia			
			AP (%)	BE (%)	EP (%)	AP (%)	BE (%)	EP (%)	AP (%)	BE (%)	EP (%)	
G1	[20]	Base	39.24 (16.93)	2.82 (4.17)	57.94 (18.85)	90.25 (3.40)	7.11 (2.44)	2.64 (1.15)	84.40 (3.88)	11.44 (2.59)	4.16 (1.83)	
		Poly	0.00 (0.00)	**0.00 (0.00)**	100.00 (0.00)	**94.54 (1.74)**	**5.20 (1.84**)	**0.27 (0.55)**	75.71 (6.30)	**7.00 (2.82)**	17.29 (5.72)	
		AR	38.11 (21.40)	5.30 (3.87)	56.59 (22.30)	85.42 (5.40)	11.47 (4.22)	3.10 (1.32)	79.18 (2.98)	16.06 (3.16)	4.75 (1.67)	
		ARX	38.32 (23.33)	4.88 (3.92)	56.80 (23.69)	85.10 (5.41)	11.67 (4.25)	3.23 (1.34)	78.96 (2.91)	16.26 (3.00)	4.78 (1.69)	
		SVR	46.89 (23.72)	6.62 (4.97)	46.49 (23.87)	86.44 (4.25)	10.64 (3.22)	2.92 (1.25)	80.90 (3.31)	14.64 (3.03)	4.46 (1.90)	
		GP	46.00 (26.35)	6.31 (3.93)	47.69 (27.28)	84.61 (5.39)	12.22 (4.16)	3.18 (1.41)	78.35 (3.63)	16.83 (3.28)	4.82 (1.60)	
		ELM	34.81 (23.43)	6.81 (4.22)	58.39 (23.88)	78.85 (4.32)	17.25 (3.18)	3.91 (1.57)	73.32 (4.41)	20.79 (3.54)	5.89 (1.73)	
		FFNN	51.88 (21.65)	3.58 (3.23)	44.54 (21.63)	82.57 (5.22)	13.73 (4.00)	3.70 (1.40)	74.60 (4.19)	19.55 (3.64)	5.84 (2.28)	
		LSTM	38.37 (23.17)	3.97 (3.72)	57.67 (24.23)	83.78 (5.33)	12.70 (4.06)	3.52 (1.47)	76.86 (3.70)	17.87 (2.73)	5.27 (2.21)	
	[21]	LWG	20.31 (17.17)	3.10 (3.81)	76.59 (16.52)	88,03 (3.46)	8.89 (2.43)	3.08 (1.20)	81.98 (3.18)	13.89 (2.24)	4.12 (2.36)	
		GE_{Φ_1}	**60.17 (34.71)**	3.58 (4.24)	**36.25 (31.24)**	91.50 (3.65)	5.82 (2.46)	2.68 (1.46)	**87.41 (3.07)**	9.04 (2.71)	**3.54 (1.90)**	
G2	[22]	$gpcLSTM_{CA}$	91.17 (8.50)	1.26 (2.08)	7.57 (8.01)	91.61 (2.03)	6.62 (1.39)	1.77 (0.74)	**87.97 (5.00)**	8.67 (2.64)	**3.36 (2.63)**	
		$gpcLSTM^{*}_{CA}$	91.02 (8.49)	**1.21 (1.97)**	7.77 (8.00)	**91.71 (2.02)**	6.55 (1.34)	**1.75 (0.77)**	87.95 (5.05)	8.69 (2.69)	3.36 (2.62)	
		GE_{Φ_2}	**91.86 (4.57)**	5.57 (3.82)	**2.57 (2.87)**	91.09 (3.63)	**5.97 (2.38)**	2.94 (1.43)	85.86 (3.59)	9.10 (2.34)	5.04 (2.17)	
G3	[22]	$gpcLSTM^{*}_{PICA}$	61.30 (20.12)	**2.92 (2.38)**	35.79 (20.23)	90.84 (3.57)	7.04 (2.57)	**2.11 (1.07)**	86.48 (3.95)	10.07 (2.66)	**3.45 (2.31)**	
		GE_{Φ_3}	**86.94 (10.27)**	5.31 (4.09)	**7.75 (9.59)**	**91.23 (3.82)**	**5.92 (2.56)**	2.85 (1.44)	**86.96 (3.25)**	**9.36 (2.48)**	3.68 (1.96)	

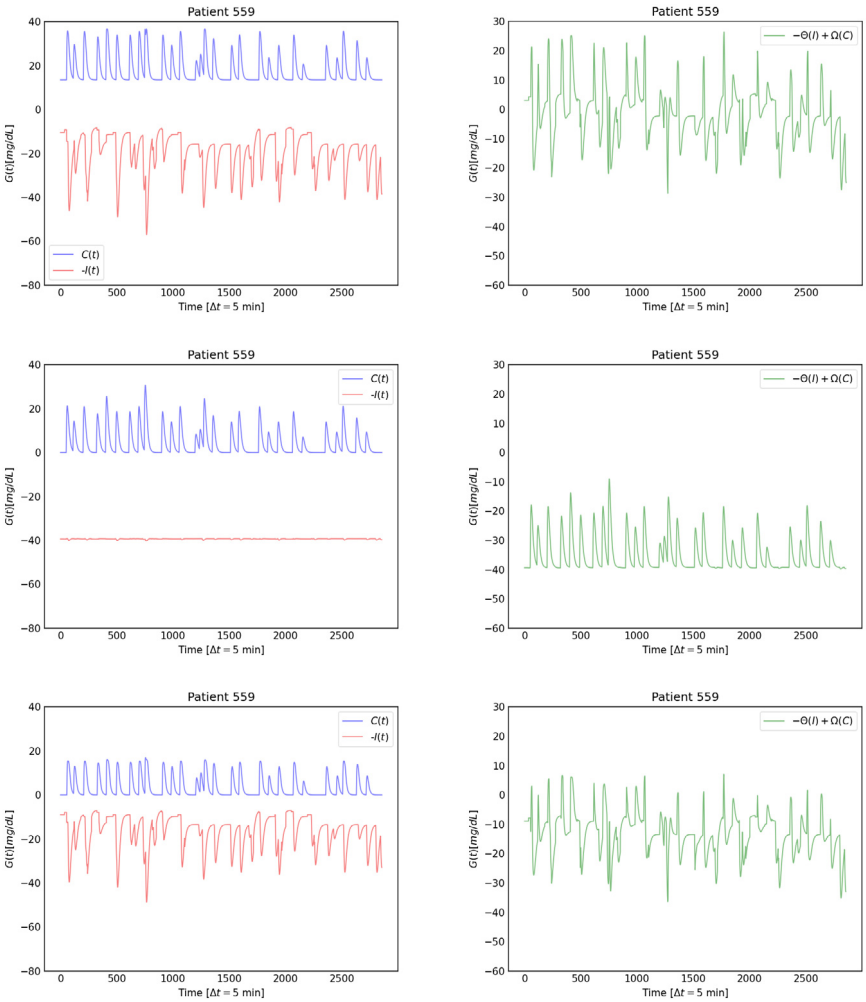

Fig. 4. Carbohydrate and insulin signals (left side) and their combined effect (right side) after the rearrangement by GE_{Φ_1} (top pane), GE_{Φ_2} (middle pane), and GE_{Φ_3} (bottom pane).

Table 6
Average numerical accuracy (with standard deviation) on the testing set computed by averaging the best glucose predictive models for each patient on the training set.

Group	Paper	Model	RMSE ([mg dl^{-1}])	MAPE (%)
G1	[20]	Ref	28.32 (2.38)	13.51 (2.72)
		Poly	57.27 (6.59)	31.09 (6.71)
		AR	20.70 (2.23)	9.62 (2.26)
		ARX	20.61 (2.20)	9.59 (2.19)
		SVR	20.10 (2.34)	**9.08 (2.12)**
		GP	**20.01 (2.33)**	9.16 (2.16)
		ELM	25.38 (1.60)	11.56 (2.43)
		FFNN	21.00 (2.24)	9.33 (2.19)
		LSTM	20.46 (2.08)	9.24 (2.10)
	[21]	LWG	22.31 (2.51)	10.96 (2.81)
		GE_{Φ_1}	21.34 (2.46)	9.91 (2.34)
G2	[22]	$gpcLSTM_{CA}$	47.70 (6.31)	**22.43 (2.76)**
		$gpcLSTM^{*}_{CA}$	47.82 (6.27)	22.47 (2.76)
		GE_{Φ_2}	**39.86 (4.11)**	22.65 (3.34)
G3	[22]	$gpcLSTM^{*}_{PICA}$	**23.50 (2.49)**	**10.46 (2.09)**
		GE_{Φ_3}	23.54 (2.32)	11.25 (2.57)

further specific considerations can be synthesized from Table 5 about the average percentages of error predictions EP in all the zones, namely, EP_i, EP_e, and EP_h, with their corresponding standard deviations. It is worth noting that the model discovered by GE_{Φ_2} shows very good performance in terms of EP_i and grants as an inheritance this capability to the model extracted through the algorithm GE_{Φ_3} which presents a very low value of EP_i. These last two algorithms present the best results in the respective groups

as concerns the error prediction in the hypoglycemic zone. In addition to lowering the error predictions in the riskiest zone, GE_{Φ_2} and GE_{Φ_3} also guarantee acceptable percentages of error predictions within euglycemic and hyperglycemic zones in their respective groups.

Further considerations related to the stability of the evolutionary algorithms can be advanced. The algorithm GE_{Φ_1} does not have optimal performance in G1, ranking as the worst and the best for the average standard deviations concerning EP_i and EP_h, respectively. Differently, the algorithms GE_{Φ_2} and on GE_{Φ_3} show the best average standard deviations for EP_i and EP_h in the respective groups.

Table 6 reports the comparison of the different approaches regarding numerical accuracy: GE_{Φ_1} has a value of RMSE (21.34 [mg dl^{-1}]) close to all the best algorithms optimizing only a numerical metric in G1. Also, compared with the algorithms by De Bois et al. focused on improving the clinical quality, i.e., $gpcLSTM_{CA}$ and $gpcLSTM^*_{CA}$ included in group G2, our analogous algorithm GE_{Φ_2} presents a better RMSE value. Finally, for the algorithms in group G3, our proposal GE_{Φ_3} has an RMSE similar to $gpcLSTM^*_{PICA}$.

All the above results confirm that the algorithm GE_{Φ_3} can represent a good compromise to save the clinical validity of the prediction without penalizing the numerical prediction accuracy, which is a very sensitive issue from a medical point of view to provide diabetic people with appropriate personalized therapy. Furthermore, as underlined above, the combination of the objective functions represents a reasonable compromise also in terms of stability.

Apart from the issue related to the clinical and numerical accuracies, for most of the above approaches, the best-performing forecasting models are obtained by leveraging different deep learning networks along with decision-level fusions of different architectures. Therefore, their prediction techniques are black boxes unable to provide explicit models. Furthermore, the fusion of the different techniques, including a deep neural network, requires a computational effort undoubtedly superior to that of a GE-based approach, not to mention the higher number of parameters to be set. For example, the LSTM model, employed as the basis of the $gpcLSTM^*_{PICA}$ approach, contains two hidden layers of 256 LSTM units, besides not providing an interpretable model. This last neural architecture exemplifies the complexity of black-box models compared to those induced by GE_{Φ_3}.

5. Conclusions

This work introduces a methodology to enhance the clinical quality of glucose predictors for diabetic patients, which also considers the numerical aspect. The improvement consists in assessing the forecasting models through an objective function that is a linear combination of clinical and numerical metrics. Specifically, the clinical metric relies on an advanced error grid analysis, i.e., PRED-EGA, which can perform accurate predictions of glucose variations preventing diabetic patients from being confused about the future glycemia evolution.

As a novel contribution, an evolutionary algorithm is exploited to improve the clinical quality of a glucose predictor evaluated through the PRED-EGA. Namely, a Grammatical Evolution algorithm has been harnessed for discovering personalized and interpretable prediction models for estimating future glucose values.

The evaluation of the discovered models over a real data set for a 30-minute time horizon has confirmed the reliability of the forecasting concerning clinical and numerical accuracy. In particular, the clinical metric has reduced the high-risk glucose forecasting errors, significantly improving confidence in predicting hypo- or hyperglycemic events. The numerical metric has contributed to keeping the absolute prediction errors low in all the regions. It is worth noting that the predicting ability of the models in the hypoglycemia region is of considerable significance from the medical side for establishing a suitable personalized diabetic therapy.

In future work, we plan to validate the proposed methodology over a broad set of clinical data to have a more and more reliable tool for minimizing the clinical risk of prediction errors in unsafe zones.

CRediT authorship contribution statement

A. Della Cioppa: Supervision, Conceptualization, Software. **I. De Falco:** Conceptualization, Methodology, Writing – review & editing. **T. Koutny:** Conceptualization, Formal analysis, Writing – review & editing. **U. Scafuri:** Methodology, Software, Investigation. **M. Ubl:** Methodology, Data curation, Validation. **E. Tarantino:** Conceptualization, Methodology, Writing – original draft.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Corresponding author is member of the editorial board. A co-author (I. De Falco) is Associate Editor of the Journal.

Data availability

The authors do not have permission to share data.

Acknowledgments

This publication was partially supported by University of West Bohemia specific research project SGS-2022-015.

References

[1] International Diabetes Federation, IDF diabetes atlas 10th ed., 2021, [Online]. Available: https://diabetesatlas.org/atlas/tenth-edition/ (Accessed 06 October 2022).
[2] World Health Organization, Diabetes fact sheet N. 312, 2013.
[3] S. Vashist, Continuous glucose monitoring systems: a review, Diagnostics 3 (4) (2013) 385–412.
[4] S. Oviedo, J. Vehí, R. Calm, J. Armengol, A review of personalized blood glucose prediction strategies for T1DM patients, Int. J. Numer. Methods Biomed. Eng. 33 (6) (2017) e2833.
[5] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, I. Chouvardal, Machine learning and data mining methods in diabetes research, Comput. Struct. Biotechnol. J. 15 (2017) 104–116.
[6] M. Rigla, G. García-Sáez, B. Pons, M.E. Hernando, Artificial intelligence methodologies and their application to diabetes, J. Diabetes Sci. Technol. 12 (2) (2018) 303–310.
[7] I. Contreras, J. Vehí, Artificial intelligence for diabetes management and decision support: literature review, J. Med. Internet Res. 20 (5) (2018) e10775.
[8] A.Z. Woldaregay, E. Årsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, G. Hartvigsen, Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes, Artif. Intell. Med. 98 (2019) 109–134.
[9] N. Tyler, P. Jacobs, Artificial intelligence in decision support systems for type 1 diabetes, Sensors 20 (3214) (2020) 1–26.
[10] V. Felizardo, N.M. Garcia, N. Pombo, I. Megdiche, Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction – A systematic literature review, Artif. Intell. Med. 118 (2021) 1–15.
[11] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell. 1 (5) (2019) 206–215.
[12] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, C. Zhong, Interpretable machine learning: Fundamental principles and 10 grand challenges, Stat. Surv. 16 (2022) 1–85.

[13] M. Du, N. Liu, X. Hu, Techniques for interpretable machine learning, Commun. ACM 63 (2020) 68–77.
[14] S.M. Lundberg, B. Nair, M.S. Vavilala, M. Horibe, M.J. Eisses, T. Adams, D.E. Liston, D.K.W. Low, S.F. Newman, J. Kim, S.I. Lee, Explainable machine-learning predictions for the prevention of hypoxaemia during surgery, Nat. Biomed. Eng. 2 (10) (2018) 749–760.
[15] W.L. Clarke, D. Cox, L. Gonder-Frederick, W. Carter, S. Pohl, Evaluating clinical accuracy of systems for self-monitoring of blood glucose, Diabetes Care 10 (1987) 622–628.
[16] J.L. Parkes, S.L. Slatin, S. Pardo, B.H. Ginsberg, A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose, Diabetes Care 23 (8) (2000) 1143–1148.
[17] D.C. Klonoff, C. Lias, R. Vigersky, W. Clarke, J.L. Parkes, D.B. Sacks, M.S. Kirkman, B. Kovatchev, E.G. Panel, The surveillance error grid, J. Diabetes Sci. Technol. 8 (4) (2014) 658–672.
[18] W.L. Clarke, S. Anderson, B. Kovatchev, Evaluating clinical accuracy of continuous glucose monitoring systems: continuous glucose-error grid analysis (CG-EGA), Curr. Diabetes Rev. 4 (3) (2008) 193–199.
[19] S. Sivananthan, V. Naumova, C.D. Man, A. Facchinetti, E. Renard, C. Cobelli, S.V. Pereverzyev, Assessment of blood glucose predictors: the prediction-error grid analysis, Diabetes Technol. Ther. 13 (8) (2011) 787–796.
[20] M. De Bois, M.A. El Yacoubi, M. Ammi, GLYFE: review and benchmark of personalized glucose predictive models in type 1 diabetes, Med. Biol. Eng. Comput. 60 (2022) 1–17.
[21] H.V. Dudukcu, M. Taskiran, T. Yildirim, Blood glucose prediction with deep neural networks using weighted decision level fusion, Biocybern. Biomed. Eng. 41 (3) (2021) 1208–1223.
[22] M. De Bois, M.A. El Yacoubi, M. Ammi, Integration of clinical criteria into the training of deep models: Application to glucose prediction for diabetic people, Smart Health 21 (2021) 100193.
[23] I. Contreras, S. Oviedo, M. Vettoretti, R. Visentin, J. Vehí, Personalized blood glucose prediction: a hybrid approach using grammatical evolution and physiological models, PLoS One 12 (11) (2017) e0187754.
[24] D. Joedicke, O. Garnica, G. Kronberger, J.M. Colmenar, S. Winkler, J.M. Velasco, S. Contador, I. Hidalgo, Analysis of the performance of genetic programming on the blood glucose level prediction challenge 2020, in: K. Bach, R. Bunescu, C. Marling, N. Wiratunga (Eds.), Proc. 5Th Int. Workshop Knowledge Discovery in Healthcare Data, Santiago de Compostela, Spain, August 29-30, 2020, pp. 147–151.
[25] S. Contador, J.M. Colmenar, O. Garnica, J.M. Velasco, J.I. Hidalgo, Blood glucose prediction using multi-objective grammatical evolution: analysis of the "agnostic" and "what-if" scenarios, Genet. Program. Evolvable Mach. 23 (2022) 161–192.
[26] M. O'Neill, C. Ryan, Grammatical evolution, IEEE Trans. Evolut. Comput. 5 (4) (2001) 349–358.
[27] P. Palumbo, S. Ditlevsen, A. Bartuzzi, A. De Gaetano, Mathematical modeling of the glucose-insulin system: A review, Math. Biosci. 44 (2013) 68–81.
[28] K. Saiti, M. Macaš, K. Štechová, P. Pithová, L. Lhotská, A review of model prediction in diabetes and of designing glucose regulators based on model predictive control for the artificial pancreas, in: Lecture Notes in Computer Science, vol. 10433, Springer, 2017, pp. 66–81.
[29] A. Nath, S. Biradar, A. Balan, R. Dey, R. Padh, Physiological models and control for type 1 diabetes mellitus: a brief review, IFAC Papers-OnLine 51 (1) (2018) 289–294.
[30] R. Visentin, E. Campos-Náñez, M. Schiavon, D. Lv, M. Vettoretti, M. Breton, B.P. Kovatchev, C.D. Man, C. Cobelli, The UVA/PADOVA type 1 diabetes simulator goes from single meal to single day, J. Diabetes Sci. Technol. 12 (2) (2018) 273–281.
[31] D. Acharya, D.K. Das, An efficient nonlinear explicit model predictive control to regulate blood glucose in type-1 diabetic patient under parametric uncertainties, Biomed. Signal Process. Control 71 (103166) (2022) 1–12.
[32] C. Novara, N.M. Pour, T. Vincent, G. Grassi, A nonlinear blind identification approach to modeling of diabetic patients, IEEE Trans. Control Syst. Technol. 24 (3) (2016) 1092–1100.
[33] J.B. Ali, T. Hamdi, N. Fnaiech, V. Di Costanzo, F. Fnaiech, J.M. Ginoux, Continuous blood glucose level prediction of type 1 diabetes based on artificial neural network, Biocybern. Biomed. Eng. 38 (4) (2018) 828–840.
[34] J. Martinsson, A. Schliep, B. Eliasson, O. Mogren, Blood glucose prediction with variance estimation using recurrent neural networks, J. Healthc. Inf. Res. 4 (2020) 1–18.
[35] C. Zecchin, A. Facchinetti, G. Sparacino, C. Cobelli, How much is short-term glucose prediction in type 1 diabetes improved by adding insulin delivery and meal content information to CGM data? A proof-of-concept study, J. Diabetes Sci. Technol. 10 (2016) 1149–1160.
[36] M. Zhang, K.B. Flores, H.T. Tran, Deep learning and regression approaches to forecasting blood glucose levels for type 1 diabetes, Biomed. Signal Process. Control. 69 (102923) (2021) 1–10.
[37] T. Yang, X. Yu, N. Ma, R. Wu, H. Li, An autonomous channel deep learning framework for blood glucose prediction, Appl. Soft Comput. 120 (2022) 108636.
[38] K. Zarkogiovanni, K. Mitsis, E. Litsa, M.-T. Arredondo, G. Fico, A. Fioravanti, K.S. Nikita, Comparative assessment of glucose prediction models for patient with type 1 diabetes mellitus applying sensors for glucose and physical activity monitoring, Med. Bio. Eng. Comput. 53 (2015) 1333–1343.
[39] A. Aliberti, I. Pupillo, S. Terna, E. Macii, S.D. Cataldo, E. Patti, A. Acquaviva, A multi-patient data-driven approach to blood glucose prediction, IEEE Access 7 (2019) 69311–69325.
[40] K. Li, J. Daniels, C. Liu, P. Herrero-Vinas, P. Georgiou, Convolutional recurrent neural networks for glucose prediction, IEEE J. Biomed. Health Inform. 24 (2020) 603–613.
[41] W. Seo, S.-W. Park, N. Kim, S.-M. Jin, S.-M. Park, A personalized blood glucose level prediction model with a fine-tuning strategy: A proof-of-concept study, Comput. Methods Programs Biomed. 211 (106424) (2021) 1–8.
[42] T. Hamdi, J.B. Ali, V. Di Costanzo, F. Fnaiech, E. Moreau, J.M. Ginoux, Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm, Biocybern. Biomed. Eng. 38 (2018) 362–372.
[43] E.I. Georga, V.C. Protopappas, D. Polyzos, D.I. Fotiadis, Evaluation of short-term predictors of glucose concentration in type1 diabetes combining feature ranking with regression models, Med. Biol. Eng. Comput. 53 (12) (2015) 1305–1318.
[44] J. Tomczak, Gaussian process regression with categorical inputs for predicting the blood glucose level, in: J. Świątek, J.M. Tomczak (Eds.), Advances in Systems Science: Proceedings of the International Conference on Systems Science (ICSS2016), 2017, pp. 98–108.
[45] I. De Falco, A. Della Cioppa, A. Marcelli, L. Stellaccio, U. Scafuri, E. Tarantino, Prediction of personalized blood glucose levels in type 1 diabetic patients using a neuroevolution approach, in: Proc. Genetic and Evolutionary Computation Conference Companion, Taipei, Taiwan, 8-10 December, 2021, pp. 1708–1716.
[46] J.M. Velasco, S. Contador, O. Garnica, I. Hidalgo, Forecasting glucose levels in patients with diabetes mellitus using semantic grammatical evolution and symbolic aggregate approximation, in: Proc. Genetic and Evolutionary Computation Conference Companion, ACM, Berlin, Germany, July 15–19, 2017, pp. 1387–1394.
[47] J.I. Hidalgo, J.M. Colmenar, G. Kronberger, S.M. Winkler, O. Garnica, J. Lanchares, Data based prediction of blood glucose concentrations using evolutionary methods, J. Med. Syst. 41 (142) (2017) 1–20.
[48] I. De Falco, A. Della Cioppa, T.T. Koutny, M. Krcma, U. Scafuri, E. Tarantino, Genetic programming-based induction of a glucose-dynamics model for telemedicine, J. Netw. Comput. Appl. 119 (2018) 1–13.
[49] I. De Falco, A. Della Cioppa, A. Giugliano, A. Marcelli, T. Koutny, M. Krcma, U. Scafuri, E. Tarantino, A genetic programming-based regression for extrapolating a blood glucose-dynamics model from interstitial glucose measurements and their first derivatives, Appl. Soft Comput. 77 (2019) 316–328.
[50] E. Georga, V. Protopappas, D. Ardigó, M. Marina, I. Zavaroni, D. Polyzos, D.I. Fotiadis, Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression, IEEE J. Biomed. Health Inform. 17 (1) (2013) 71–81.
[51] K. Zarkogiovanni, E. Litsa, A. Vazeou, K.S. Nikita, Personalized glucose-insulin metabolism model based on self-organizing maps for patients with type 1 diabetes mellitus, in: Proceedings 13rd IEEE Symposium on Bioinformatics and Bioengineering, Nov. 10-13, 2013, Chania, Greece, 2013, pp. 1–3.
[52] M. Cescon, R. Johansson, E. Renard, Subspace-based linear multi-step predictors in type 1 diabetes mellitus, Biomed. Signal Process. Control 22 (2015) 99–110.
[53] K.S. Eljil, G. Qadah, M. Pasquier, Predicting hypoglycemia in diabetic patients using time-sensitive artificial neural network, Int. J. Healthcare Inf. Syst. Inform. 11 (4) (2016) 70–88.
[54] R.H.K. Karim, I. Vassány, I. Kósa, After-meal blood glucose level prediction using an absorption model for neural network training, Comput. Biol. Med. 125 (103956) (2020) 1–13.
[55] I.J. Hidalgo, J.M. Colmenar, J. L.Risco-Martin, A. Cuesta-Infante, E. Maqueda, M. Botella, Modeling glycemia in humans by means of Grammatical Evolution, Appl. Soft Comput. 20 (2014) 40–53.
[56] J.M. Velasco, O. Garnica, S. Contador, J. Lanchares, E. Maqueda, M. Botella, J.I. Hidalgo, Enhancing grammatical evolution through data augmentation: application to blood glucose forecasting, in: Applications of Evolutionary Computation, Part I - Lecture Notes in Computer Science, vol. 10199, Springer, 2017, pp. 142–157.
[57] J.M. Velasco, O. Garnica, J. Lanchares, M. Botella, J.I. Hidalgo, Combining data augmentation, EDAs and grammatical evolution for blood glucose forecasting, Memetic Comput. 10 (3) (2018) 267–277.

[58] N.C. Lourenço, J. Comenar, J.I. Hidalgo, O. Garnica, Structured grammatical evolution for glucose prediction in diabetic patients, in: Proc. Genetic and Evolutionary Computation Conference Companion, 2017, pp. 1250–1257.
[59] J.I. Hidalgo, M. Botella, J.M. Velasco, O. Garnica, C. Cervigón, R. Martínez, A. Aramendi, E. Maqueda, J. Lanchares, Glucose forecasting combining markov chain based enrichment of data, random grammatical evolution and bagging, Appl. Soft Comput. 88 (2020) 1–12.
[60] I. De Falco, A. Della Cioppa, T. Koutny, M. Krcma, U. Scafuri, E. Tarantino, A grammatical evolution approach for estimating blood glucose levels, in: Proc. 11th IEEE Global Communications Conf. - Int. Workshop on AI-Driven Smart Healthcare (AIdSH), Taipei, Taiwan, 8-10 December, 2020, pp. 1–6.
[61] I. De Falco, A. Della Cioppa, T. Koutny, U. Scafuri, E. Tarantino, M. Ubl, Grammatical evolution-based approach for extracting interpretable glucose-dynamics models, in: Proc. IEEE Conference on ICT Solutions for EHealth, IEEE, Athens, Greece, September 5-8, 2021, pp. 91–95.
[62] R. Hovorka, V. Canonico, L.J. Chassin, U. Haueter, M. Massi-Benedetti, M. Orsini Federici, T.R. Pieber, H.C. Schaller, L. Schaupp, T. Vering, M.E. Wilinska, Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes, Physiol. Meas. 25 (2004) 905–920.
[63] R. Hovorka, J.K. Powrie, G.D. Smith, P.H. Sönksen, E.R. Carson, R.H. Jones, Five-compartment model of insulin kinetics and its use to investigate action of chloroquine in NIDDM, Am. J. Physiol. 265 (1993) 162–1752.
[64] G. Livesey, P.D. Wilson, J.R. Dainty, J.C. Brown, R.M. Faulks, M.A. Roe, T.A. Newman, J. Eagles, F.A. Mellon, R.H. Greenwood, Simultaneous time-varying systemic appearance of oral and hepatic glucose in adults monitored with stable isotopes, Am. J. Physiol. 275 (1998) 717–728.
[65] S. Sampath, V. Naumova, C.D. Man, A. facchinetti, E. Renard, C. Cobelli, S. Pereverzyev, Assessment of blood glucose predictors: The prediction-error grid analysis, Diabetes Technol. Ther. 13 (8) (2011) 787–796.
[66] S. Del Favero, A. Facchinetti, C. Cobelli, A glucose-specific metric to assess predictors and identify models, IEEE Trans. Biomed. Eng. 59 (5) (2012) 1281–1290.
[67] J. Ni, R. Drieberg, P. Rockett, The use of an analytic quotient operator in genetic programming, IEEE Trans. Evol. Comput. 17 (1) (2013) 146–152.
[68] C. Marling, R. Bunescu, The ohiot1DM dataset for blood glucose level prediction, in: K. Bach, R. Bunescu, O. Farri, A. Guo, S. Hasan, Z. Ibrahim, C. Marling, J. Raffa, J. Rubin, H. Wu (Eds.), 3rd International Workshop on Knowledge Discovery in Healthcare Data, KDH, 2018, pp. 60–63.
[69] M. Sevil, M. Rashid, I. Hajizadeh, M. Park, L. Quinn, A. Cinar, Physical activity and psychological stress detection and assessment of their effects on glucose concentration predictions in diabetes management, IEEE Trans. Biomed. Eng. 68 (7) (2021) 2251–2260.
[70] M. Fenton, J. McDermott, D. Fagan, S. Forstenlechner, E. Hemberg, M. O'Neill, PonyGE2: grammatical evolution in python, in: Proc. Genetic and Evolutionary Computation Conference Companion, Springer, Berlin, Germany, July 15-19, 2017, pp. 1194–1201.