Computer Methods and Programs in Biomedicine 221 (2022) 106862

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Generation of post-meal insulin correction boluses in type 1 diabetes simulation models for *in-silico* clinical trials: More realistic scenarios obtained using a decision tree approach

N. Camerlingo[a], M. Vettoretti[a], S. Del Favero[a], A. Facchinetti[a], P. Choudhary[b], G. Sparacino[a,*], on behalf of the Hypo-RESOLVE Consortium

[a] Department of Information Engineering, University of Padova, Via G. Gradenigo 6B, Padova 35131, Italy
[b] Department of Diabetes, Leicester Diabetes Centre, University of Leicester, Gwendolen Rd, Leicester LE5 4PW, United Kingdom

ARTICLE INFO

Article history:
Received 4 February 2022
Revised 19 April 2022
Accepted 7 May 2022

Keywords:
Type 1 diabetes
Insulin bolus
Decision tree
lASSO
In-silico clinical trial

ABSTRACT

Background and objective: In type 1 diabetes (T1D) research, *in-silico* clinical trials (ISCTs) notably facilitate the design/testing of new therapies. Published simulation tools embed mathematical models of blood glucose (BG) and insulin dynamics, continuous glucose monitoring (CGM) sensors, and insulin treatments, but lack a realistic description of some aspects of patient lifestyle impacting on glucose control. Specifically, to effectively simulate insulin correction boluses, required to treat post-meal hyperglycemia (BG > 180 mg/dL), the timing of the bolus may be influenced by subjects' behavioral attitudes. In this work, we develop an easily interpretable model of the variability of correction bolus timing observed in real data, and embed it into a popular simulation tool for ISCTs.

Methods: Using data collected in 196 adults with T1D monitored in free-living conditions, we trained a decision tree (DT) model to classify whether a correction bolus is injected in a future time window, based on predictors collected back in time, related to CGM data, previous insulin boluses and subject's characteristics. The performance was compared to that of a logistic regression classifier with LASSO regularization (LC), trained on the same dataset. After validation, the DT was embedded within a popular T1D simulation tool and an ISCT was performed to compare the simulated correction boluses against those observed in a subset of data not used for model training.

Results: The DT provided better classification performance (accuracy: 0.792, sensitivity: 0.430, specificity: 0.878, precision: 0.455) than the LC and presented good interpretability. The most predictive features were related to CGM (and its temporal variations), time since the last insulin bolus, and time of the day. The correction boluses simulated by the DT, after implementation in the simulation tool, showed a good agreement with real-world data.

Conclusions: The DT developed in this work represents a simple set of rules to mimic the same timing of correction boluses observed on real data. The inclusion of the model in simulation tools allows investigators to perform ISCTs that more realistically represent the patient behavior in taking correction boluses and the post-prandial BG response. In the future, more complex models can be investigated.

© 2022 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Type 1 diabetes (T1D) is a chronic metabolic disease characterized by the almost total absence of endogenous insulin, a hormone responsible for blood glucose (BG) regulation [1]. Individuals with T1D require a healthy diet, physical exercise, and a lifelong exogenous insulin therapy to maintain their BG within the recommended target range (i.e., 70–180 mg/dL). Specifically, to counter-

Abbreviations: AUROC, area under the receiver operating characteristic curve; BG, Blood glucose; BMI, body mass index; BW, body weight; CART, classification and regression tree; CF, correction factor; CGM, continuous glucose monitoring; CHAID, chi-squared automatic interaction detection; CV, cross validation; DT, decision tree; GDI, Gini's diversity index; ISCT, *in-silico* clinical trial; LASSO, least absolute shrinkage and selector operator; LC, logistic classifier; PrS, pruning set; REP, reduced error pruning; SAP, sensor augmenting pump; SVM, support vector machine; T1D, Type 1 diabetes; T1D-PDS, Type-1 diabetes Patient Decision Simulator; TeS, test set; TrS, training set.

* Corresponding author.
E-mail address: gianni@dei.unipd.it (G. Sparacino).

https://doi.org/10.1016/j.cmpb.2022.106862
0169-2607/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

balance the rise in BG produced by food intake, individuals take insulin boluses administered at mealtimes (also referred to as meal boluses), while the so-called correction boluses are recommended to be injected to correct, or at least mitigate, the risk of hyperglycemia (i.e., BG $>$ 180 mg/dL) [2].

Nowadays, new personalized T1D therapies can be easily designed /tested thanks to the availability of simulation tools relying on mathematical models of glucose-insulin system dynamics, and allow the performance of *in-silico* clinical trials (ISCTs) [3,4]. ISCTs enable large-scale testing at limited costs, without implicating any risk for real subjects. As such, in the past 15 years of T1D research, ISCTs have accelerated the development of new treatments [5,6] and drugs [7–9], and have facilitated the design of clinical studies [10–12].

While numerous literature simulation tools effectively capture the T1D individual physiology [13–16], the mathematical description of therapy-related behaviors has been rarely investigated so far. A first attempt to simulate these aspects, thus enabling more realistic ISCTs, was the prototype behavioral model embedded in the T1D Patient Decision Simulator (T1D-PDS) of Vettoretti et al. [17]. This model allowed describing some aspects of behaviors and lifestyle that have a large impact on glucose control, such as the variability in meal time and amount, the behavior in consuming rescue carbohydrates to avoid low BG levels, and the errors in timing of the meal bolus time and in carbohydrate counting [18]. Though the T1D-PDS was proven effective in augmenting the reliability of ISCTs, [19,20] its behavioral model did leave some room for improvement. In recent works by our research group, we enhanced the reliability of the T1D-PDS by developing and embedding new mathematical models of meal amount and timing variability [21] and carb-counting error [22].

Another margin for improving the T1D-PDS of Vettoretti et al. [17] is given by the very simplistic model of insulin correction bolus timing, which only considers empirical thresholds on glucose levels and trend, thus not reflecting the much more complex patients' behavior in injecting correction boluses.

Several guidelines exist to recommend effective correction bolus timing. For example, American Diabetes Association standards of diabetes care [23] recommend patients should take a correction bolus whenever BG concentration is above 180 mg/dL, and at least 2 h have passed since the last meal. With the recent advent of continuous glucose monitoring (CGM) sensors, minimally-invasive devices providing glucose readings and rate-of-change (usually displayed as an arrow) almost continuously [19], more complex guidelines have been proposed. Aleppo et al. [24] recommend modifying the insulin dose by a fixed insulin amount, depending on the displayed arrow, the time passed since the last meal (i.e., from 2 to 4 h, or more than 4 h), and a subject-specific parameter that is the correction factor (CF). To give a practical example, if at least 4 h have passed since the mealtime, the sensor shows one up arrow, and the individual's CF is $\geq$ 75, then the guidelines of Aleppo et al. suggest increasing the correction bolus amount of 1 insulin unit.

Despite the availability of the above-mentioned guidelines, individuals with T1D often do not (or cannot) strictly adhere to these recommendations. In fact, the timing of a correction bolus injection may be decided by individuals based on different conditions, not only related to glucose levels (e.g., CGM readings and glucose trend) and previous insulin injections, but also to the specific moment of the day (e.g., if it is during daylight or night), the daily activities, the fear for hypoglycemia, hyperglycemia symptoms, and other factors. Identifying the circumstances that prompt individuals to be more likely to take a correction bolus could be useful to design (or optimize) new personalized insulin therapies.

The aim of this work is twofold: i) to develop a new data-driven mathematical model of post-meal correction bolus timing, and ii) to embed the new model into the T1D-PDS, to enhance its capability of simulating the patients' behavior in injecting correction boluses. The model of correction bolus timing is developed using a decision tree (DT) model, since it is remarkably easy to read and interpret, and it allows mimicking the human decision-making process.

More in detail, first we use data collected in T1D individuals monitored in free-living conditions to train a DT classifier, able to determine whether a post-meal insulin correction bolus is injected in future time windows, based on predictors collected back in time. The classification performance of the new DT model is evaluated against those of a logistic regression classifier with LASSO regularization (LC), used as benchmark for comparison. Then, we embed the new DT model into the T1D-PDS, and we perform an ISCT to compare the resulting simulated correction boluses against those collected in a real-world dataset.

2. Methods

2.1. Dataset

In this work we considered two different studies, for sake of readability labelled as A and B.

Study A [25,26] involved 30 adults with T1D, recruited from six medical centers in Europe and USA, undergoing sensor augmented pump (SAP) therapy. On mean±std, participants were 40.3 $\pm$ 12.8 years old, had diabetes duration of 21.4 $\pm$ 7.47 years, glycated haemoglobin (HbA1c) of 7.3 $\pm$ 0.7% and body mass index (BMI) of 25.5 $\pm$ 4.3 kg/m^2. Participants were monitored in free-living conditions for 9.8 $\pm$ 7.1 days, using the Dexcom G4 Platinum CGM sensor (Dexcom Inc., San Diego, CA, USA), the Accu-Chek insulin pump (Roche Diagnostics, Mannheim, Germany), and the DiAs [27], a smartphone-based medical platform which allows tracking many variables, such as carbohydrates intakes (distinguishing between meals and rescue carbohydrates) and insulin boluses (distinguishing between meal boluses and correction boluses).

Study B [28] involved 166 T1D adult patients undergoing SAP therapy. On mean±std, participants were 41.9 $\pm$ 12.3 years old, had diabetes duration of 20.2 $\pm$ 12.2 years, HbA1c of 8.3 $\pm$ 0.5% and BMI of 27.4 $\pm$ 4.4 kg/m^2. Participants were monitored in free-living conditions for 477.6 $\pm$ 208.1 days using the MiniMed Paradigm REAL-Time System (Medtronic Inc., Minneapolis, MN, USA) [29], a device integrating an insulin pump with a CGM sensor. This system embedded a bolus wizard calculator, allowing to store information on size and timing of meals and insulin boluses (distinguishing between meal boluses and correction boluses).

2.2. Data pre-processing

A preliminary analysis of the raw datasets was performed to identify and remove data that appears to have been inaccurately reported. For example, in the raw datasets a small number of correction boluses resulted performed with low CGM values (i.e., $<$120 mg/dL). Such boluses, which in all likelihood were meal boluses of unreported meals, were discarded from the subsequent analysis. Few insulin boluses of very small amount ($<$0.4 U) were also discarded, since they were associated to cannula refill [30]. In addition, when two correction boluses of the same size were reported in less than 20 min, a duplication of the same information was assumed to have been recorded and the latter of the two was ignored.

The same pipeline described in Camerlingo et al. [21] was implemented to label the meal intakes as breakfast, lunch, dinner, or snack. Finally, the days with few data gathered (i.e., those with less than 3 meal intakes) were not considered for our analysis.

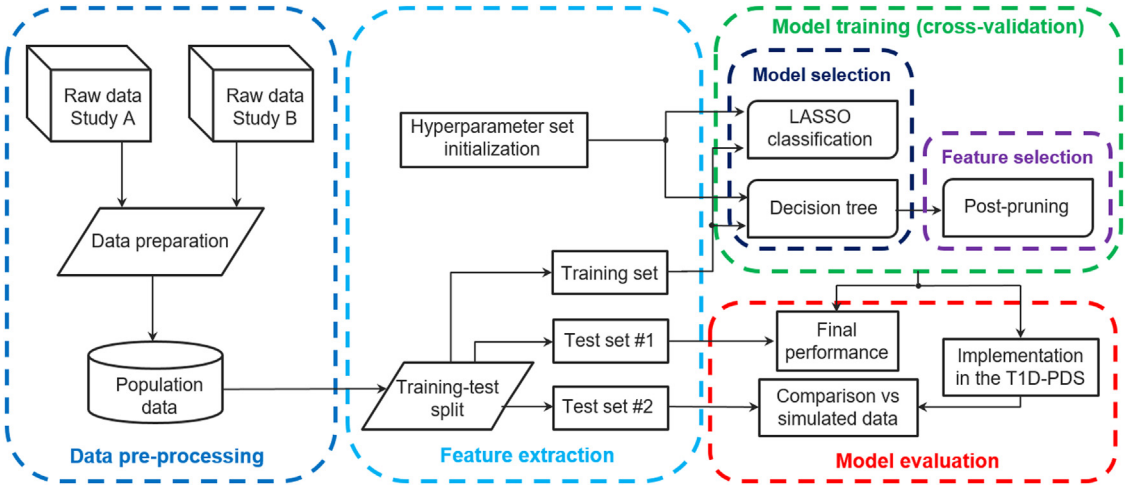

Fig. 1. Design of the analysis.

The final dataset included a total of 49,995 days, 296,685 meals (including snacks), and 61,654 insulin correction boluses.

The pipeline implemented to model the behavior of subjects with T1D in injecting post-meal correction boluses is illustrated in Fig. 1.

2.3. Feature extraction

For each meal recorded in the dataset, we considered a post-meal window starting 30 min after the meal and ending after 7 h, if no other meal has been consumed up to that moment, otherwise the window was truncated 30 min before the following meal. For the latest collected meal, if closer than 7 h to the tail of the data collection period, the longest window with duration multiple of 30 min was considered. Each post-meal window was then split in shorter consecutive windows of 30-min duration, from now on referred to as observation windows. An observation window was labelled with "1", if at least one correction bolus was registered inside the window, otherwise the observation window was labelled with "0". The total number of observation windows was 460,602: 29,799 labelled as "1" and 430,802 labelled as "0".

Then, for each observation window, 12 independent variables were extracted both from current and previous windows and from patients' demographic data, as possible predictors of the label. Specifically, we considered the following features**:**

- Patient's age.
- Patient's body weight (BW).
- Patient's CF.
- Mean glucose trend direction in the previous 30-min, computed, at each sample time, as the weighted average of the last 5 CGM samples, and then discretized as an integer number ranging in [−3, 3], to mimic the different arrow directions displayed by CGM sensors.
- Current CGM, obtained by averaging the CGM samples of the observation window.
- ΔCGM, obtained as difference of current CGM and the average of the previous 30 min CGM samples.
- Mealtime ΔCGM, obtained as difference of current CGM and the average of the previous CGM samples since the last meal.
- Time spent in hyperglycemia, computed as percent of CGM samples >180 mg/dL since the last meal intake.
- Type of last meal intake, a boolean variable equal to 1 if the last meal was labelled a snack, 0 otherwise.
- Time from the last insulin bolus, including both meal boluses and correction boluses.
- type of last insulin bolus, a boolean variable equal to 1 if the last insulin bolus was a correction bolus, 0 if it was a meal bolus.
- Time of the day, a categorical variable equal to 1, 2, 3, 4 if the first sample in the observation window was collected in the time interval 5:00 am–11:55 am, 12:00 am–5:55 pm, 6:00 pm–11:55 pm, 00:00 am–4:55 am, respectively.

The final dataset was randomly divided into a 70% training set (TrS) and 30% test set (TeS), maintaining the same proportions of the labels: 301,460 (93.5%) "0", 20,957 (6.50%) "1" in the TrS, and 129,197 (93.5%) "0", 8982 (6.50%) "1" in the TeS. The TeS was further halved into TeS1 and TeS2: the former will be used to evaluate the model performance, while the latter will be used as comparison against simulated data.

Since the number of "0" is much greater than the number of "1", the label's frequency results imbalanced. In order to mitigate bias issues, TrS data were weighted to penalize the misclassification of the minority class, according to the following weighting scheme:

$$w_k = \frac{N}{Kn_k} \tag{2.3.1}$$

where $N = 322,417$ was the total number of observations in the TrS, $K = 2$ was the total number of classes, and n_k was the number of observations in the class k, thus obtaining $w_{k=0} = 0.535$ and $w_{k=1} = 7.692$. This approach, often referred to as cost-sensitive learning, is preferred to deal with class-imbalance problems as it allows using all the available training data, thus avoiding loss of information, without degrading the learning speed, compared to other commonly used approaches [31].

Moreover, to alleviate the subject-specific features' variability, a z-score standardization was performed on the continuous features using their mean and their standard deviation in the TrS [32].

2.4. Classification model selection for DT and LC

Two supervised learning techniques were implemented to predict the dependent variable based on the independent variables collected in each window: DT, and LC. These models have been widely applied in the literature to approach similar problems [21,33,34].

The DT is a highly interpretable commonly used approach for classification problems [35,36]. A DT may be grown through several algorithms, including "classification and regression tree" (CART) [37], "chi-squared automatic interaction detection" (CHAID)

[38], and extensive CHAID [39], tested in this work. While the former chooses the split predictor based on the maximization of a split criterion gain among all the possible splits, the other two perform chi-square tests of independence between each predictor and response (CHAID), or also each pair of predictors and response (extensive CHAID). For the CART algorithm, we tested two different split criteria: Gini's diversity index (GDI) and cross-entropy. Other hyperparameters, including the minimum number of observations in the leaf node and the maximum number of branch nodes, were tuned via grid search analysis. In particular, the minimum number of observations in the leaf node ranged within (10, 50), with a step of 10, while the maximum number of branch nodes ranged within (5, 40), with a step of 5. This search space was chosen to penalize too deep trees, prone to overfitting and more difficult to read.

The LC is a binary classification model, where the dependent variable is assumed to be a linear combination of the independent variables, transformed by the logistic function. Thus, it was implemented as a generalized linear model in which the likelihood of the outcome was assumed to be a Bernoulli distribution [40]. To avoid overfitting and deal with multicollinearity [41], we resorted to the least absolute shrinkage and selection operator (LASSO) regularization approach, that essentially consists in adding an L1 penalty term (i.e., a bias) to the objective function, thus achieving a lower variance on the TeS [42]. This leads to an overall improvement in the accuracy of classification [43]. The amount of regularization applied is controlled by the hyperparameter λ: the larger its value, the higher the incidence of the penalty term in the whole objective function and, therefore, the lower the model coefficients. In this work, λ ranged within (10^{-5}, 10^{-1}), with a step of 10^{-3}.

All the hyperparameters for each model were tuned by applying a 20-fold cross validation (CV) over the TrS, to achieve the highest area under the receiver operating characteristic curve (AUROC). The AUROC is commonly employed in classification problems, since it quantifies to what extent the model distinguishes between classes: the closer the AUROC is to one, the better is the discriminatory power of the model.

2.5. Feature selection

Feature selection algorithms were implemented to eliminate the redundant features not relevant for the target classification task.

Regarding the DT, on one hand large DTs are often inaccurate because of a too large variance (too much sample sensitivity), whereas on the other hand DTs with too few leaves are inaccurate because of a large bias (not enough flexibility). To find a suitable bias-variance trade-off and to perform feature selection, the bottom-up reduced error pruning (REP) approach was implemented [44,45]. Specifically, a 30% TrS data was randomly extracted as a pruning set (PrS), maintaining the same proportions of the labels, while the remaining 70% was used to grow the maximal expanded tree. For each internal node of this tree, the algorithm compared the number of classification errors made on the PrS when the node was turned into a leaf and associated with its most popular class. If this simplified tree exhibited a non-worse performance than the original one, quantified in terms of AUROC, pruning was confirmed. This operation was repeated on the simplified tree until no more branches could be pruned. In this way, REP provided the small version of the most accurate subtree with respect to the pruning set [46].

For the LC, feature selection is automatically obtained by the LASSO regularization approach. Indeed, the LASSO model sets as many coefficients as possible to zero, unless they are important to drive the predictions right.

Table 1
Performance metrics used for model selection.

Metric	Expression
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Sensitivity (recall)	$\frac{TN}{TN+FP}$
Specificity	$\frac{TP}{TP+FN}$
Precision	$\frac{TP}{TP+FP}$

TP, true positive, window correctly classified with a correction bolus; TN, true negative, window correctly classified without a correction bolus; FP, false positive, window wrongly classified with a correction bolus; FN, false negative, window wrongly classified without a correction bolus.

Lastly, DT, and LC models containing their respective subsets of most predictive features were trained on the whole TrS, and their performance were computed on TeS1, in terms of sensitivity, i.e., the true positive rate, specificity, i.e., the true negative rate, accuracy, i.e., the proportion of correct predictions, precision, i.e., the positive predictive value, and AUROC. These metrics were summarized in Table 1. In this computation, we considered as true positive both the windows correctly classified as "1", and those windows misclassified as "1" that were located immediately before or immediately after a window labelled as "1" in the original dataset. The choice of accepting a prediction in the windows surrounding those with a correction bolus was based on the consideration that sometimes, based on the specific circumstances, a correction bolus can be delayed (e.g., a subject who is driving, first needs to stop before injecting a correction bolus) or anticipated (e.g., an individual who is leaving the house may decide to anticipate the correction bolus, thus avoiding the burden of injecting it in presence of friends/colleagues).

2.6. Implementation in the T1D-PDS

Among the previous models, the one providing the best performance was embedded into the T1D-PDS, published in Vettoretti et al. [17]. An ISCT involving 100 virtual subjects monitored for 7 days was designed.

For each virtual subject, one breakfast, one lunch, and one dinner were always triggered during the day, at times selected by extracting random samples from the distributions of breakfast time, time between breakfast-lunch, and time between lunch-dinner, derived in Camerlingo et al. [21]. Similarly, meal amounts were sampled from the distributions of breakfast amount, lunch amount, and dinner amount. Snacks were triggered according to a SVM model, applied every 3 h [21]. Finally, the duration of main meals and snacks was set to 15 min and 5 min, respectively.

Both main meals and snacks were associated with insulin meal boluses, whose doses was set according to the standard formula [47], i.e., by taking into account the patient's estimate of the carbohydrates content of the meal, the mealtime CGM value, and patient-specific therapy-related factors. In particular, the estimated meal content was simulated by implementing the nonlinear model of carb-counting error developed in Roversi et al. [22].

Predictors of future correction boluses were collected in real-time, and the final model was applied every 30 min, from 30 min after each meal to the following meal, up to 7 h. Then, if the model predicted a correction bolus in the following half hour, a correction bolus was triggered at a time randomly selected, with uniform probability, within the time window. The correction bolus amount was calculated by subtracting the target blood glucose from the current blood glucose and dividing by CF.

Once incorporated the model into the T1D-PDS, to demonstrate the reliability of its realizations, several insulin-related outcomes were computed and compared against the same metrics computed over TeS2 data. Specifically, assessment metrics were:

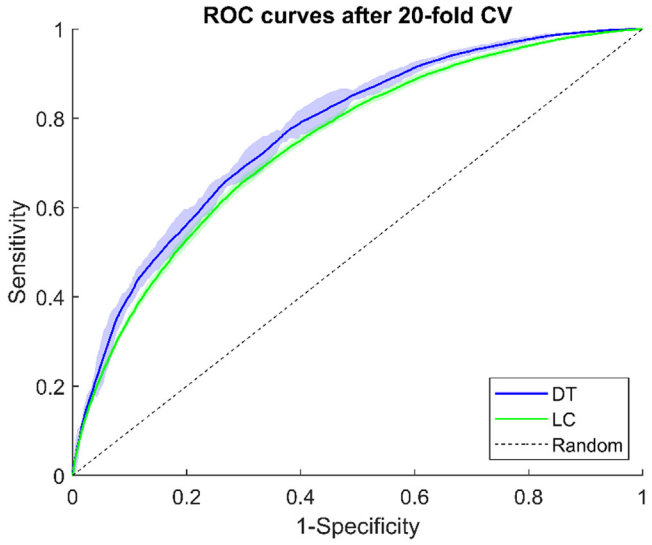

Fig. 2. ROC curves of the models selected in 20-fold CV for each learning technique: DT (blue), and LC (green), reported as mean [25–75th percentiles], compared against the random classifier (dashed black).

Table 2
Performance of the models in CV, using the entire set of features (columns 2–3) and on the TeS1, using the selected subset of features (columns 4–5).

Performance metrics	All features in 20-fold CV		Selected features on the TeS1	
	DT	LC	DT	LC
ACC	0.800 [0.795–0.802]	0.791 [0.785–0.800]	0.790	0.792
SEN	0.385 [0.375–0.397]	0.411 [0.389–0.447]	0.501	0.430
SPE	0.895 [0.889–0.900]	0.877 [0.867–0.892]	0.861	0.878
PRE	0.459 [0.443–0.469]	0.435 [0.419–0.446]	0.467	0.455
AUROC	0.750 [0.743–0.757]	0.743 [0.738–0.749]	0.774	0.757

CGM value at correction bolus injection time [mg/dL], time between consecutive correction boluses, $\Delta t_{CB\text{-}CB}$ [min], time between a correction bolus and the previous meal, $\Delta t_{m\text{-}CB}$ [min], number of correction boluses per day, #CB/day, total amount of insulin injected per day as correction boluses, $Insulin_{CB}$/day [U], total amount of insulin injected per day as meal bolus plus correction bolus, $Insulin_{CB+MB}$/day [U]. In addition, the average numbers of CB per day and CB per meal over the real and the virtual populations were compared.

3. Results

The DT trained with the CART algorithm, with a minimum leaf size of 100 points and a maximum number of splits of 30 showed the highest AUROC among those tested in CV, equal to 0.750 [0.743–0.757], on median [25–75th percentiles]. The highest performance for LC was achieved with regularization factor $\lambda = 2.4 \times 10^{-3}$, which provided AUROC equal to 0.743 [0.738–0.749], on median [25–75th percentiles].

The 20-fold CV average ROC curves of DT and LC are reported in Fig. 2, together with their [25–75th percentiles] intervals. The models generate probability estimates for each prediction. In this work, we tuned the threshold probability to maximize the models' accuracy. In the second column of Table 2, the resulting values of accuracy, sensitivity, specificity, and precision of the models are reported on median [25–75th percentiles]. In CV, the DT performed

Table 3
Coefficients of the LC model using LASSO regularization.

Candidate predictors	Model coefficient
Current CGM	**0.4351**
Daytime = 2	**0.3954**
Intercept	**−0.3915**
Time from last bolus	**0.3433**
Delta CGM	**0.2575**
Delta CGM from mealtime	**0.2548**
Daytime = 1	**0.2437**
Daytime = 3	**−0.1587**
BW	**0.1059**
Time in hyperglycemia	0
Previous correction bolus	0
Patient's CF	0
Glucose trend direction	0
Meal type	0
Patient's age	0

better than the LC, with higher median values of accuracy (0.800 for the DT vs 0.791 for the LC), specificity (0.895 for the DT vs 0.877 for the LC), and precision (0.459 for the DT vs 0.435 for the LC), and a slightly lower value of sensitivity (0.385 for the DT vs 0.411 for the LC).

Then, the models were trained on the whole TrS. Fig. 3 reports the resulting DT, that involves the following features: current CGM, time from the last insulin bolus, ΔCGM, mealtime ΔCGM, and time of the day.

The coefficients of the LC are reported in Table 3. Those set to zero were related to the features deemed poorly predictive by the LASSO regularization approach, while the subset of most predictive features included: current CGM, time of the day, time from the last insulin bolus, ΔCGM, mealtime ΔCGM, and patient's BW.

The final ROC curves are reported in Fig. 4. The thresholds on the posterior probability are: 0.591 for the DT, and 0.613 for the LC, and are marked by colored dots in Fig. 4. The corresponding performance metrics evaluated on TeS1 are reported in the third column of Table 2.

On the TeS1, the DT performed slightly better than the LC, with an AUROC of 0.774, compared to 0.757 of the LC. Despite a comparable accuracy (0.790 for the DT vs 0.792 for the LC), the DT exhibits a higher sensitivity (0.501 vs 0.430) and a higher precision (0.467 vs 0.455), but a lower specificity (0.861 vs 0.878).

Furthermore, the DT resulted easy to understand and visualize, providing a simple set of rules to determine the most suitable moment to perform a correction bolus, based on previously collected predictors. Thus, implementing the DT into the T1D-PDS would be preferable.

After embedding the DT into the T1D-PDS, a total of 2554 meals and 784 post-meal correction boluses were generated. In order to assess whether the model could capture the variability observed in real-world data, in Fig. 5 the distributions of CGM value at correction bolus injection time (panel a), time between consecutive correction boluses, $\Delta t_{CB\text{-}CB}$ (panel b), time between a correction bolus and the previous meal, $\Delta t_{m\text{-}CB}$ (panel c), number of correction boluses per day, #CB/day (panel d), total amount of insulin injected per day as correction boluses, $Insulin_{CB}$/day (panel e), total amount of insulin injected per day as meal bolus plus correction bolus, $Insulin_{CB+MB}$/day (panel f) are shown through boxplot representation for both real data of TeS2 (label "Data") and simulated data (label "Sim"). The metrics present similar distributions in real and simulated datasets. In Table 4, we report the median [25–75th percentiles] of these metrics, calculated on real data (second column) and simulated data (third column).

Lastly, the average numbers of correction bolus per day were 1.14 for the real population and 1.12 for the virtual population,

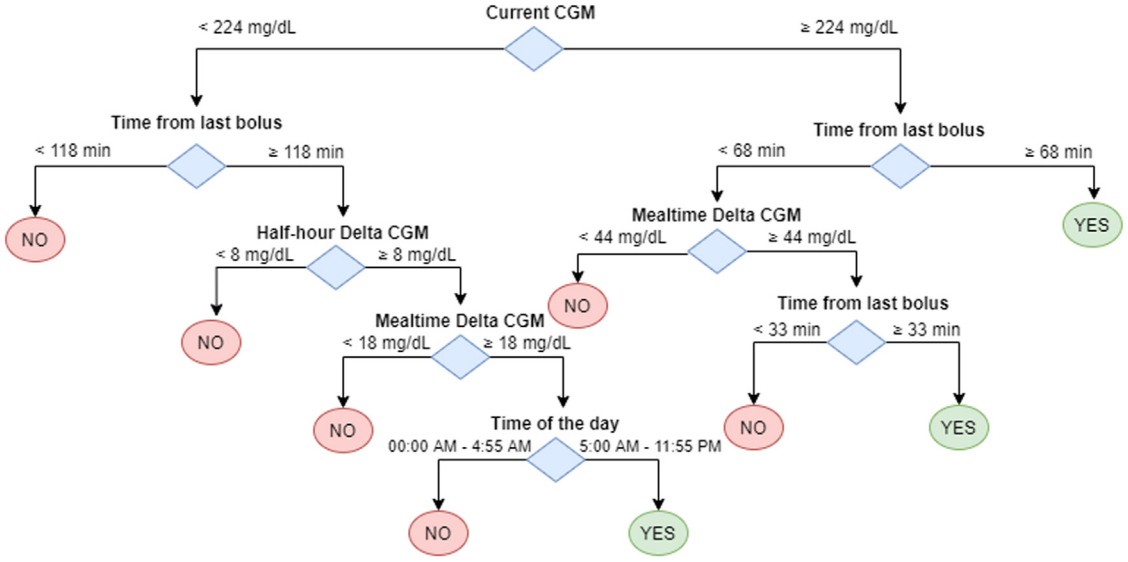

Fig. 3. Final DT trained on the entire TR.

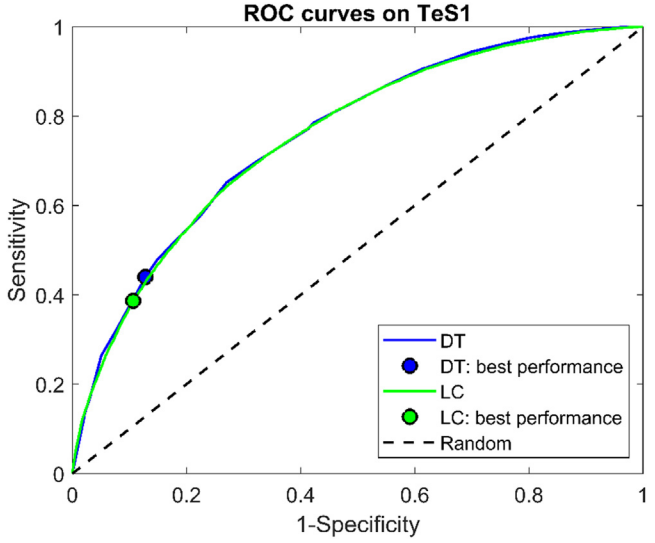

Fig. 4. Final ROC curves of the models trained on the whole TR: DT (blue curve), and LC (green curve). Colored dots indicates the points in which the maximum accuracy is reached.

Table 4
Insulin outcomes computed on data of real population vs virtual population.

Metric	Real population	Virtual population
CGM at correction bolus timing [mg/dL]	242 [210–284]	249 [225–295]
Time between consecutive correction boluses [min]	125 [115–145]	114 [100–130]
Time between meal and correction bolus [min]	154 [90–205]	138 [114–175]
Number of correction bolus per day	1 [0–2]	1 [0–2]
Insulin amount injected as correction bolus [U]	3.7 [2.0–6.7]	3.5 [0.0–6.8]
Insulin amount injected as meal bolus and correction bolus [U]	24.3 [16.4–33.5]	24.2 [18.7–31.5]

while the average number of correction bolus per meal were 0.22 for the real population and 0.36 for the virtual population.

4. Discussion

Existing simulation tools to perform ISCTs in T1D research lack a realistic description of some individuals' behavior in performing daily therapeutic tasks, that may impact on glucose control. In particular, the T1D-PDS of Vettoretti et al. [17] embeds a very simplistic and empiric behavioral module that did leave room for improvement. While we have already focused on the enhancement of meal amount and timing variability [21] and carb-counting error [22] models, in this work we developed a realistic model of insulin correction bolus timing.

Specifically, by leveraging two datasets collected in free-living conditions, we implemented two different machine learning classifiers to describe the timing of performing post-meal insulin correction bolus injections. We developed a DT to classify whether a correction bolus is injected in a future time window, based on predictors collected back in time, that could likely influence the injection of a correction bolus and, at the same time, can be reliably handled by the ad-hoc models embedded in the currently available simulation tools. These predictors are related to CGM data, previous insulin boluses and subject's characteristics. The model was assessed both in CV and on a separate test set, and it was compared against a LC with LASSO approach. The DT showed better performance compared to the LC, and resulted highly interpretable, providing a bunch of easily understandable rules, with a logical clinical meaning. The final DT involved the following features: current CGM, time from the last insulin bolus, half-hour ΔCGM, mealtime ΔCGM, and time of the day. Specifically, in case of CGM levels below 224 mg/dL, a correction bolus is injected only during the day (5:00 AM–11:55 PM), if at least 118 min have passed since the last insulin bolus, the CGM increased of at least 8 mg/dL in the last half-hour and it increased of at least 18 mg/dL since the last meal. On the contrary, in case of CGM levels equal to or above 224 mg/dL, a correction bolus is injected either if at least 68 min have passed since the last insulin bolus, or if the CGM increased of more than 44 mg/dL since the last meal and at least 33 min have passed since the last insulin bolus.

On the test set, the DT showed an accuracy of 0.790, a sensitivity of 0.501, a specificity of 0.861, a precision of 0.467 and an AUROC of 0.774. The overall performance, despite not optimal, are satisfactory for the purpose of the models developed in this work,

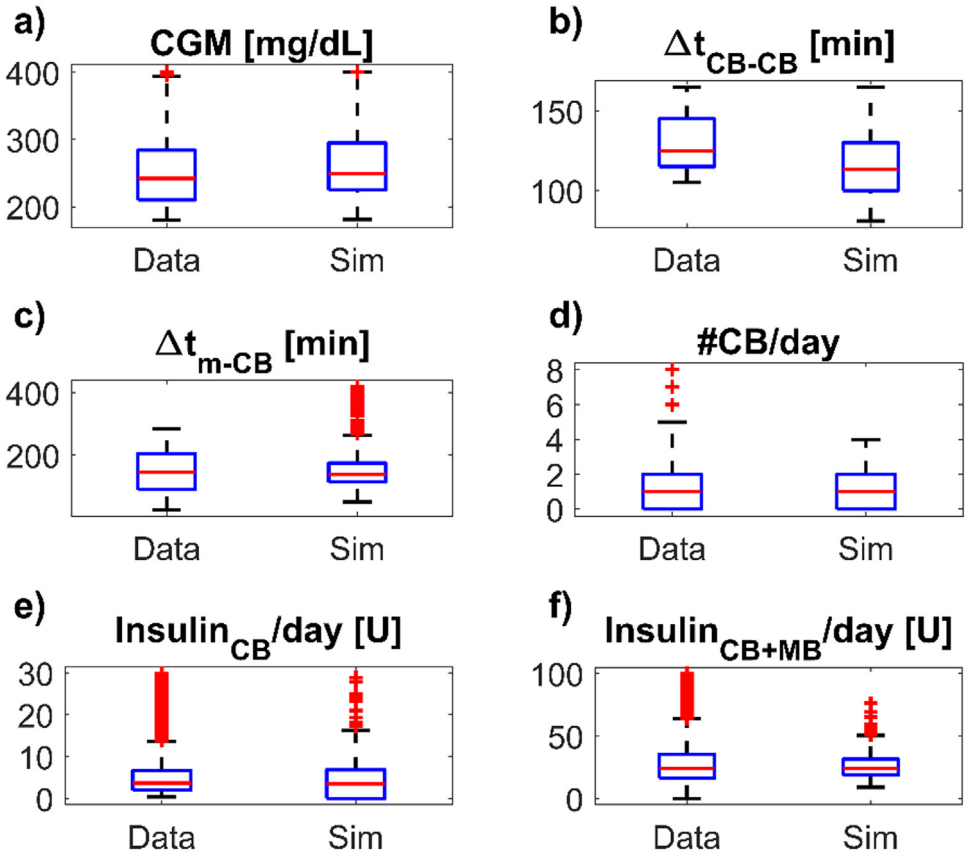

Fig. 5. Boxplot representation of the distributions of CGM during correction bolus (panel a), time between consecutive correction boluses, ΔtCB-CB (panel b), time between a correction bolus and the previous meal, Δtm-CB (panel c), number of correction boluses per day, #CB/day (panel d), total amount of insulin injected per day as correction boluses, InsulinCB/day (panel e), total amount of insulin injected per day as meal bolus plus correction bolus, InsulinCB+MB/day (panel f), obtained on real data (label "Real") and simulated data (label "Sim"). The red horizontal line represents median, the blue box markes the interquartile range, dashed black lines are the whiskers and the red stars indicate outliers.

i.e., realistically simulate the patients' behavior in post-meal correction bolus timing.

The new DT was incorporated into the recent T1D-PDS and its reliability was assessed by comparing the simulated correction boluses of 100 virtual subjects to a portion of the correction boluses collected in the study used in this work, but not used for training the model. The comparison highlighted good agreement between the metrics calculated on real and on simulated data.

In the future, more complex models can be investigated to possibly achieve better performance metrics, such as support vector machine, random forest, and neural networks, whose interpretability could be achieved by recently developed reverse-engineering algorithms, e.g., SHAP [48]. Future developments could also include the refinement of the model to capture the temporal patterns of patients' behavior at different time scales (e.g., working days vs weekend, different seasons, etc.). Moreover, the availability of a much larger dataset would allow both to identify other subject-specific covariates (for example, linked to the education level, lifestyle, and possible comorbidities), and to develop new behavioral models (for example, to describe the insulin blousing behavior, the probability of a missed bolus, and the responsiveness to CGM alarms/alerts). This can be achieved either by implementing the same methodology presented therein, or by resorting to reinforcement-learning-based techniques to learn sequential decision-making tasks (as performed in [49]).

In conclusion, the work carried out in this work allowed to further enhance the reliability of the T1D-PDS in performing more insightful ISCTs. According to the "Hypo-RESOLVE" project objectives [50], the enhanced simulator will be used to quantify the impact of different behaviors in performing daily therapeutic tasks on glucose control.

Declaration of Competing Interest

NC, MV, AF, GS and SDF declare no conflicts of interest to disclose. PC has received personal fees from Medtronic, Abbott, Dexcom, Insulet, Novo Nordisk, Lilly, Sanofi.

Acknowledgments

This study is part of the Hypo-RESOLVE project. The project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No. 777460. The JU receives support from the European Union's Horizon 2020 research and innovation program and EFPIA and T1D Exchange, JDRF, International Diabetes Federation (IDF), The Leona M. and Harry B. Helmsley Charitable Trust.

References

[1] D. Daneman, Type 1 diabetes, Lancet 367 (2006) 847–858, doi:10.1016/S0140-6736(06)68341-4.
[2] I.H. de Boer, B. Kestenbaum, T.C. Rue, M.W. Steffes, P.A. Cleary, M.E. Molitch, J.M. Lachin, N.S. Weiss, J.D. Brunzell, Insulin therapy, hyperglycemia, and hypertension in type 1 diabetes mellitus, Arch. Intern. Med. 168 (2008) 1867–1873, doi:10.1001/ARCHINTERNMED.2008.2.
[3] M. Viceconti, A. Henney, E. Morley-Fletcher, *In silico* clinical trials: how computer simulation will transform the biomedical industry, Int. J. Clin. Trials. 3 (2016) 37, doi:10.18203/2349-3259.ijct20161408.
[4] F. Pappalardo, G. Russo, F.M. Tshinanu, M. Viceconti, *silico* clinical trials: concepts and early adoptions, Brief. Bioinform. 20 (2019) 1699–1708, doi:10.1093/BIB/BBY043.

[5] N. Camerlingo, M. Vettoretti, S. Del Favero, G. Cappon, G. Sparacino, A. Facchinetti, A real-time continuous glucose monitoring-based algorithm to trigger hypotreatments to prevent/mitigate hypoglycemic events, Diabetes Technol. Ther. 21 (2019) 644–655, doi:10.1089/dia.2019.0139.
[6] M.D. Breton, R. Hinzmann, E. Campos-Nañez, S. Riddle, M. Schoemaker, G. Schmelzeisen-Redeker, Analysis of the accuracy and performance of a continuous glucose monitoring sensor prototype: an *in-silico* study using the UVA/PADOVA type 1 diabetes simulator, J. Diabetes Sci. Technol. 11 (2017) 545–552, doi:10.1177/1932296816680633.
[7] K. Rehman, S.M. Munawar, M.S.H. Akash, M.A. Buabeid, T.A. Chohan, M. Tariq, K. Jabeen, E.S.A. Arafa, Hesperidin improves insulin resistance via down-regulation of inflammatory responses: biochemical analysis and *in silico* validation, PLoS ONE 15 (2020) e0227637, doi:10.1371/journal.pone.0227637.
[8] M. Schiavon, R. Visentin, C. Giegerich, J. Sieber, C. Dalla Man, C. Cobelli, T. Klabunde, *In silico* head-to-head comparison of insulin glargine 300 U/mL and insulin degludec 100 U/mL in type 1 diabetes, Diabetes Technol. Ther. (2020) 2020.0027 dia., doi:10.1089/dia.2020.0027.
[9] A. Dasgupta, G.K. Bandyopadhyay, I. Ray, K. Bandyopadhyay, N. Chowdhury, R.K. De, S.K. Mahata, Catestatin improves insulin sensitivity by attenuating endoplasmic reticulum stress: *in vivo* and *in silico* validation, Comput. Struct. Biotechnol. J. 18 (2020) 464–481, doi:10.1016/j.csbj.2020.02.005.
[10] B.P. Kovatchev, M. Breton, C. Dalla Man, C. Cobelli, *In silico* preclinical trials: a proof of concept in closed-loop control of type 1 diabetes, J. Diabetes Sci. Technol. 3 (2009) 44–55, doi:10.1177/193229680900300106.
[11] P. Herrero, J. Bondia, O. Adewuyi, P. Pesl, M. El-Sharkawy, M. Reddy, C. Toumazou, N. Oliver, P. Georgiou, Enhancing automatic closed-loop glucose control in type 1 diabetes with an adaptive meal bolus calculator – *in silico* evaluation under intra-day variability, Comput. Methods Progr. Biomed. 146 (2017) 125–131, doi:10.1016/j.cmpb.2017.05.010.
[12] A. Bajard, S. Chabaud, C. Cornu, A.C. Castellan, S. Malik, P. Kurbatova, V. Volpert, N. Eymard, B. Kassai, P. Nony, An *in silico* approach helped to identify the best experimental design, population, and outcome for future randomized clinical trials, J. Clin. Epidemiol. 69 (2016) 125–136, doi:10.1016/j.jclinepi.2015.06.024.
[13] R. Visentin, E. Campos-Náñez, M. Schiavon, D. Lv, M. Vettoretti, M. Breton, B.P. Kovatchev, C. Dalla Man, C. Cobelli, The UVA/padova type 1 diabetes simulator goes from single meal to single day, J. Diabetes Sci. Technol. 12 (2018) 273–281, doi:10.1177/1932296818757747.
[14] R. Hovorka, F. Shojaee-Moradie, P.V. Carroll, L.J. Chassin, I.J. Gowrie, N.C. Jackson, R.S. Tudor, A. Margot Umpleby, R.H. Jones, Partitioning glucose distribution/transport, disposal, and endogenous production during IVGTT, Am. J. Physiol. Endocrinol. Metab. 282 (2002) 992–1007, doi:10.1152/ajpendo.00304.2001.
[15] M. Rashid, S. Samadi, M. Sevil, I. Hajizadeh, P. Kolodziej, N. Hobbs, Z. Maloney, R. Brandt, J. Feng, M. Park, L. Quinn, A. Cinar, Simulation software for assessment of nonlinear and adaptive multivariable control algorithms: glucose–insulin dynamics in type 1 diabetes, Comput. Chem. Eng. 130 (2019) 106565, doi:10.1016/j.compchemeng.2019.106565.
[16] P.G. Fabietti, V. Canonico, M.O. Federici, M.M. Benedetti, E. Sarti, Control oriented model of insulin and glucose dynamics in type 1 diabetics, Med. Biol. Eng. Comput. 44 (2006) 69–78, doi:10.1007/s11517-005-0012-2.
[17] M. Vettoretti, A. Facchinetti, G. Sparacino, C. Cobelli, Type-1 diabetes patient decision simulator for *in silico* testing safety and effectiveness of insulin treatments, IEEE Trans. Biomed. Eng. 65 (2018) 1281–1290, doi:10.1109/TBME.2017.2746340.
[18] A.S. Brazeau, H. Mircescu, K. Desjardins, C. Leroux, I. Strychar, J.M. Ekoé, R. Rabasa-Lhoret, Carbohydrate counting accuracy and blood glucose variability in adults with type 1 diabetes, Diabetes Res. Clin. Pract. 99 (2013) 19–23, doi:10.1016/j.diabres.2012.10.024.
[19] G. Cappon, M. Vettoretti, G. Sparacino, A. Facchinetti, Continuous glucose monitoring sensors for diabetes management: a review of technologies and applications, Diabetes Metab. J. 43 (2019) 383–397, doi:10.4093/dmj.2019.0121.
[20] J. Garcia-Tirado, C. Zuluaga-Bedoya, M.D. Breton, Identifiability analysis of three control-oriented models for use in artificial pancreas systems, J. Diabetes Sci. Technol. 12 (2018) 937–952, doi:10.1177/1932296818788873.
[21] N. Camerlingo, M. Vettoretti, S. Del Favero, A. Facchinetti, G. Sparacino, Mathematical models of meal amount and timing variability with implementation in the type-1 diabetes patient decision simulator, J. Diabetes Sci. Technol. 15 (2020) 346–359, doi:10.1177/1932296820952123.
[22] C. Roversi, M. Vettoretti, S. Del Favero, A. Facchinetti, G. Sparacino, Modeling carbohydrate counting error in type 1 diabetes management, Diabetes Technol. Ther. 22 (2020) 749–759, doi:10.1089/dia.2019.0502.
[23] American Diabetes Association, Diabetes Care in the Hospital: Standards of Medical Care in Diabetes—2021, Diabetes Care 1 (2021) S211–S220, doi:10.2337/dc21-S015.
[24] G. Aleppo, L.M. Laffel, A.J. Ahmann, I.B. Hirsch, D.F. Kruger, A. Peters, R.S. Weinstock, D.R. Harris, A practical approach to using trend arrows on the Dexcom G5 CGM system for the management of adults with diabetes, J. Endocr. Soc. 1 (2017) 1445–1460, doi:10.1210/JS.2017-00388.
[25] S.M. Anderson, D. Raghinaru, J.E. Pinsker, F. Boscari, E. Renard, B.A. Buckingham, R. Nimri, F.J. Doyle, S.A. Brown, P. Keith-Hynes, M.D. Breton, D. Chernavvsky, W.C. Bevier, P.K. Bradley, D. Bruttomesso, S. Del Favero, R. Calore, C. Cobelli, A. Avogaro, A. Farret, J. Place, T.T. Ly, S. Shanmugham, M. Phillip, E. Dassau, I.S. Dasanayake, C. Kollman, J.W. Lum, R.W. Beck, B. Kovatchev, Multinational home use of closed-loop control is safe and effective, Diabetes Care 39 (2016) 1143–1150, doi:10.2337/DC15-2468.
[26] B. Kovatchev, C. Peiyao, S.M. Anderson, J.E. Pinsker, F. Boscari, B.A. Buckingham, F.J. Doyle, H.K. Hood, M.D. Breton, D. Chernavvsky, W.C. Bevier, P.K. Bradley, D. Bruttomesso, S. Del Favero, R. Calore, C. Cobelli, A. Avogaro, T.T. Ly, S. Shanmugham, E. Dassau, C. Kollman, J.W. Lum, R.W. Beck, Feasibility of long-term closed-loop control: a multicenter 6-month trial of 24/7 automated insulin delivery, Diabetes Technol. Ther. 19 (2017) 18–24, doi:10.1089/dia.2016.0333.
[27] P. Keith-Hynes, S. Guerlain, B. Mize, C. Hughes-Karvetski, M. Khan, M. McElwee-Malloy, B.P. Kovatchev, DiAs user interface: a patient-centric interface for mobile artificial pancreas systems, J. Diabetes. Sci. Technol. 7 (2013) 1416–1426, doi:10.1177/193229681300700602.
[28] R.M. Bergenstal, W.V. Tamborlane, A. Ahmann, J.B. Buse, G. Dailey, S.N. Davis, C. Joyce, T. Peoples, B.A. Perkins, J.B. Welsh, S.M. Willi, M.A. Wood, Effectiveness of sensor-augmented insulin-pump therapy in type 1 diabetes, New Engl. J. Med. 363 (2010) 311–320, doi:10.1056/NEJMOA1002853.
[29] J. Mastrototaro, S. Lee, The integrated minimed paradigm real-time insulin pump and glucose monitoring system: implications for improved patient outcomes, Diabetes Technol. Ther. 11 (2009) S37–S43, doi:10.1089/DIA.2008.0134.
[30] S.W. Ponder, J.S. Skyler, D.F. Kruger, D. Matheson, B.W. Brown, Unexplained hyperglycemia in continuous subcutaneous insulin infusion, Diabetes Educ. 34 (2008) 327–333, doi:10.1177/0145721708315682.
[31] G.M. Weiss, Mining with rarity, ACM SIGKDD Explor. Newsl. 6 (2004) 7–19, doi:10.1145/1007730.1007734.
[32] D.C. Luor, A comparative assessment of data standardization on support vector machine for classification problems, Intell. Data Anal. 19 (2015) 529–546, doi:10.3233/IDA-150730.
[33] M.M. Ghiasi, S. Zendehboudi, A.A. Mohsenipour, Decision tree-based diagnosis of coronary artery disease: CART model, Comput. Methods Progr. Biomed. 192 (2020) 105400, doi:10.1016/J.CMPB.2020.105400.
[34] S. Cui, D. Wang, Y. Wang, P.W. Yu, Y. Jin, An improved support vector machine-based diabetic readmission prediction, Comput. Methods Progr. Biomed. 166 (2018) 123–135, doi:10.1016/J.CMPB.2018.10.012.
[35] R.Z. Liu, B. Fang, H.W. Luo, Automatic decision support by rule exhaustion decision tree algorithm, Int. Conf. Wavelet Anal. Pattern Recognit. 11 (2016) 25–30, doi:10.1109/ICWAPR.2016.7731623.
[36] A. Iyer, S. Jeyalatha, R. Sumbaly, Diagnosis of diabetes using classification mining techniques, Int. J. Data Min. Knowl. Manag. Process. 5 (2015) 01–14, doi:10.5121/ijdkp.2015.5101.
[37] N. Speybroeck, Classification and regression trees, Int. J. Public Health 571 (2011) 57 (2011) 243–246, doi:10.1007/S00038-011-0315-Z.
[38] W.Y. Loh, Y.S. Shih, Split selection methods for classification trees, Stat. Sin. 7 (1997) 815–840 https://www.jstor.org/stable/24306157. (accessed October 19, 2021).
[39] W.Y. Loh, Regression trees with unbiased variable selection and interaction detection, Stat. Sin. 12 (2002) 361–386 https://www.jstor.org/stable/24306967. (accessed October 19, 2021).
[40] Y. Yang, M. Loog, A benchmark and comparison of active learning for logistic regression, Pattern Recognit. 83 (2018) 401–415, doi:10.1016/J.PATCOG.2018.06.004.
[41] C.F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J.R.G. Marquéz, B. Gruber, B. Lafourcade, P.J. Leitão, T. Münkemüller, C. McClean, P.E. Osborne, B. Reineking, B. Schröder, A.K. Skidmore, D. Zurell, S. Lautenbach, Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, Ecography 36 (2013) 27–46 (Cop.), doi:10.1111/J.1600-0587.2012.07348.X.
[42] D. Urda, L. Franco, J.M. Jerez, Classification of high dimensional data using LASSO ensembles, in: Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI) 2017, 1, 2018, pp. 1–7, doi:10.1109/SSCI.2017.8280875.
[43] T. Hastie, R. Tibshirani, J.H. Friedman, in: The elements of statistical learning: data mining, inference, and prediction, springer, New York, 2009, pp. 1–758, doi:10.1007/978-0-387-21606-5.
[44] C. Iorio, M. Aria, A. D'Ambrosio, R. Siciliano, Informative trees by visual pruning, Expert Syst. Appl. 127 (2019) 228–240, doi:10.1016/J.ESWA.2019.03.018.
[45] W.N.H.W. Mohamed, M.N.M. Salleh, A.H. Omar, A comparative study of reduced error pruning method in decision tree algorithms, in: Proceedings of the IEEE International Conference on Control System, Computing and Engineering ICCSCE 2012, 2012, pp. 392–397, doi:10.1109/ICCSCE.2012.6487177.
[46] F. Esposito, D. Malerba, G. Semeraro, A comparative analysis of methods for pruning decision trees, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 476–491, doi:10.1109/34.589207.
[47] G. Cappon, F. Marturano, M. Vettoretti, A. Facchinetti, G. Sparacino, *In silico* assessment of literature insulin bolus calculation methods accounting for glucose rate of change, J. Diabetes Sci. Technol. 13 (2019) 103–110, doi:10.1177/1932296818777524.
[48] S.M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2019) https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.
[49] T. Zhu, K. Li, P. Herrero, P. Georgiou, Basal glucose control in type 1 diabetes using deep reinforcement learning: an *in silico* validation, IEEE J. Biomed. Health Inform. 25 (2021) 1223–1232, doi:10.1109/JBHI.2020.3014556.
[50] B.E. de Galan, R.J. McCrimmon, M. Ibberson, S.R. Heller, P. Choudhary, F. Pouwer, J. Speight, J. Carlton, T.R. Pieber, M. Rosilio, C.J. Tack, M. Müllenborn, Reducing the burden of hypoglycaemia in people with diabetes through increased understanding: design of the hypoglycaemia redefining solutions for better lives (hypo-resolve) project, Diabet. Med. 37 (2020) 1066–1073, doi:10.1111/DME.14240.