

Evaluation of short-term predictors of glucose concentration in type 1 diabetes combining feature ranking with regression models

Eleni I. Georga¹ · Vasilios C. Protopappas¹ · Demosthenes Polyzos² · Dimitrios I. Fotiadis¹

Received: 31 March 2014 / Accepted: 27 February 2015 / Published online: 15 March 2015
© International Federation for Medical and Biological Engineering 2015

Abstract Glucose concentration in type 1 diabetes is a function of biological and environmental factors which present high inter-patient variability. The objective of this study is to evaluate a number of features, which are extracted from medical and lifestyle self-monitoring data, with respect to their ability to predict the short-term subcutaneous (s.c.) glucose concentration of an individual. Random forests (RF) and RReliefF algorithms are first employed to rank the candidate feature set. Then, a forward selection procedure follows to build a glucose predictive model, where features are sequentially added to it in decreasing order of importance. Predictions are performed using support vector regression or Gaussian processes. The proposed method is validated on a dataset of 15 type diabetics in real-life conditions. The s.c. glucose profile along with time of the day and plasma insulin concentration are systematically highly ranked, while the effect of food intake and physical activity varies considerably among patients. Moreover, the average prediction error converges in less than $d/2$ iterations (d is the number of features). Our results suggest that RF and RReliefF can find the most informative features and can be successfully used to customize the input of glucose models.

Keywords Type 1 diabetes · Feature selection · Glucose level prediction · Machine learning

✉ Dimitrios I. Fotiadis
fotiadis@cc.uoi.gr

¹ Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, 45110 Ioannina, Greece

² Department of Mechanical Engineering and Aeronautics, University of Patras, 26500 Patras, Greece

1 Introduction

Daily management of type 1 diabetes at its core can be viewed as a feedback loop where patients adjust the insulin regime based primarily on their real-time blood glucose measurements and secondarily on their overall lifestyle context (e.g., meals, physical activities, stress) [3]. Patients on multiple insulin injection therapy could maintain complete glycemic control, if it were not for the increased risk of hypoglycemia [7]. The technological progress in continuous glucose monitoring (CGM) and in continuous subcutaneous insulin infusion (CSII) has contributed to a more practical and safe therapy scheme [4, 29, 30]. In particular, sensor-augmented insulin pump (SAP) therapy has been shown to reduce glycemic variability. Until researchers close the loop in insulin delivery, the patient, as the main actor in this process, should continually reason the effect of insulin intake and lifestyle on his/her glucose metabolism.

Prediction algorithms of subcutaneous (s.c.) glucose concentration have the potential to further advance insulin-treated diabetes management either in open or in semi-closed loop conditions [6, 11, 22, 45]. Initial approaches to this problem, which were based on autoregressive (AR) or autoregressive moving average models (ARMA) of the CGM time series either with constant [16, 33] or with recursively identified [12, 37] parameters, had sufficiently accurate short-term (15 and 30 min) predictive capacity. The combination of a recurrent neural network (RNN) with compartmental models of plasma insulin concentration and carbohydrates absorption was proposed in [27, 44]. Zecchin et al. [47] demonstrated that feed-forward and jump neural networks [46] exploiting not only the past CGM data but also meal information allow for improved accuracy when compared to [28, 37] over 30-min horizon. In addition, the

inclusion of real physical activity data in glucose predictive models has recently emerged, which is very important considering the prominent effect of exercise on glucose concentration. In our previous work [18], we proposed an individualized predictive model relying on support vector regression (SVR) of multiple input variables concerning the recent s.c. glucose concentration profile, the effect of food and insulin intake, the energy expenditure in physical activities and the time of the day. We demonstrated that these additional inputs not only result in better short-term predictions but also make feasible the predictions for longer horizons (60 and 120 min). Similarly, a patient-specific recursive ARMA model with exogenous inputs (ARMAX) from a multi-sensor body monitor outperformed a univariate model as applied to type 2 diabetics [13] and allowed the accurate prediction of hypoglycemic events in type 1 diabetics [41, 42].

The existent inter- and intra-patient variability in type 1 diabetes implies the individualization of the predictive models and their continuous adaptation to both biological and environmental changes as well [1, 8]. For instance, the fusion of real-time adaptive models (RNN and AR) resulted in 100 % prediction accuracy of hypoglycemic events for patients under SAP therapy during everyday living conditions [8]. This need can also be partially met by performing a periodic patient-specific training process. A complementary procedure to adaptive learning can be considered the individualized evaluation of the short-term predictors of glucose concentration and the subsequent refinement of the model's input [43]. In [18], we predefined a high-dimensional feature set in an attempt to represent spatial and temporal input–output dependencies. As expected, we found out considerable inter-patient deviations in the hyper-parameters of the SVR regarding the same input, which means that it can be further customized.

In this study, we propose feature ranking as a preprocessing step in the construction of patient-specific predictive models of the short-term s.c. glucose concentration in type 1 diabetes. Two well-established feature ranking algorithms suitable for regression problems, i.e., random forests (RF) and RReliefF, are employed for assessing the set of features defined in [18] separately for each patient. RF is a prediction technique that incorporates feature ranking as part of the training process [5], while RRelief is a pure feature filtering algorithm based on the nearest neighbors approach [34]. Their main advantages which render them appropriate for the specific application are as follows: (1) the sensitivity to informative features as well as to the correlations among them, (2) the absence of assumptions about the (non)linearity of the underlying function and (3) the low computational complexity [34, 39]. The generality and effectiveness of the result of feature ranking is demonstrated with respect to the performance of a nonlinear regression model for the estimation of glucose concentration. Herein, we choose SVR and

Gaussian processes (GP) kernel-based methods as prediction tools since they have been shown to perform equally well over the full range of glucose values in our previously developed dataset [19]. In this context, the top-ranked features obtained per individual are examined and a clinical interpretation of the results is attempted. To our knowledge, this is the first work which examines the concurrent and cumulative impact of the most important predictors of the short-term daily glucose dynamics in a type 1 diabetic individual (i.e., meals, insulin therapy, physical activities and the glucose signal itself) with the aid of feature ranking.

2 Materials and methods

2.1 Subjects

A short-term observational study was carried out in two centers (Parma University Hospital, Parma and University Hospital Motol, Prague) as part of a European Union co-funded research project named METABO [17]. The study was approved by the ethics committees of each hospital. Fifteen type 1 diabetic patients, following multiple-dose insulin therapy and without significant micro- and macro-vascular complications, were monitored from 5 to 22 days (average 12.5 ± 4.6) in free-living conditions. All subjects provided written informed consent before enrollment. Patients wore the Guardian Real-Time CGM system (Medtronic Minimed Inc.) which reports an average s.c. glucose value every 5 min. In addition, they were equipped with the SenseWear Armband (BodyMedia Inc.) physical activity monitor which computes energy expenditure every 1 min. Information on food intake (i.e., type of food, serving sizes and time) and insulin regime (i.e., type of insulin, injection dosage and time) was also recorded on a daily basis using a specially designed paper diary. The amount of carbohydrates for each meal was post-analyzed by a dietitian. Table 1 presents the baseline characteristics of the patients and some descriptive statistics of their CGM data.

2.2 Dataset construction

A separate dataset $D^{(s)} = \{(x^i, y_i) | i = 1, \dots, N_s\}$ is constructed for each subject s . Each sample associates the input vector $x^i \in \mathbb{R}^d$ at time t_i with the observed glucose concentration y_i at time $t_i + l$, where l is the prediction horizon. The feature set $F = \{F_1, \dots, F_d\}$ is defined with respect to the present time (i.e., t) and the prediction horizon l as follows:

- $F_1 = hh$: the hour of the day associated with time t .
- $F_{2-8} = \{gl(t - 30), \dots, gl(t - 5), gl(t)\}$: s.c. glucose measurements within the last 30 min.

Table 1 Description of the dataset

Patient baseline characteristics	
Gender	3 F/12 M
Age	40.3 ± 13.5
BMI	25.2 ± 2.9
HbA1C	7.1 ± 1.2
Descriptive statistics of glucose dataset	
Average s.c. glucose concentration (mg/dl)	145.9 ± 22.8
Min s.c. glucose concentration (mg/dl)	49.7 ± 10.1
Max s.c. glucose concentration (mg/dl)	333.8 ± 48.7
% of hypoglycemic values	0.05 ± 0.04
% of hyperglycemic values	0.25 ± 0.14

Data are mean ± standard deviation values

A glucose concentration value ≤70 mg/dl is defined as hypoglycemic

A glucose concentration value ≥180 mg/dl is defined as hyperglycemic

- $F_{9-15} = \{Ra(t + l - 30), \dots, Ra(t + l - 5), Ra(t + l)\}$: rate of appearance of meal-derived glucose into plasma within the time interval $[t + l - 30, t + l]$ [25].
- $F_{16-21} = \{SRa(t + l - 75), \dots, SRa(t + l - 15), SRa(t + l)\}$: total glucose inserted into plasma calculated cumulatively every 15 min over the last 90 min with respect to $t + l$, where $SRa(t + l - (5 - i)15) = \sum_{\tau=t+l-90}^{t+l-(75-15i)} Ra(\tau)$ for $i = 0, \dots, 5$.
- $F_{22-28} = \{I_p(t + l - 30), \dots, I_p(t + l - 5), I_p(t + l)\}$: plasma insulin concentration within the time interval $[t + l - 30, t + l]$ [40].
- $F_{29-46} = \{SEE(t - 170), \dots, SEE(t - 10), SEE(t)\}$: energy expenditure calculated cumulatively every 10 min over the last 3 h where $SEE(t - (17 - i)10) = \sum_{\tau=t-180}^{t-(170-10i)} EE(\tau)$ for $i = 0, \dots, 17$. The term EE expresses the instantaneous (i.e., per minute) energy expenditure estimated by the physical activity monitor.

In particular, a new sample is added into $D^{(s)}$ for each time instance in the glucose time series of subject s for which all $d = 46$ features can be defined and the value of glucose concentration l min ahead is available. The size of dataset, N_s , depends mainly on the length of the observation period for each patient, and ideally, the time difference between two consecutive samples in $D^{(s)}$ is equal to the sampling period of the glucose time series, i.e., 5 min. Nevertheless, the existence of gaps in the sensor data reduces N_s . In addition, all samples (x^i, y_i) for which an event (i.e., food intake, insulin intake, moderate or intense exercise) exists within the time interval $[t_i, t_i + l]$ are excluded from $D^{(s)}$ since they do not represent a rational mapping between the configured input and the output. This also ensures that for all samples in $D^{(s)}$, the upcoming values of Ra and I_p within $[t_i, t_i + l]$ have

been computed based only on the insulin and meal recordings until t_i .

2.3 Feature ranking

The feature set $F = \{F_j\}$, with $j = 1, \dots, d$, is evaluated individually for each subject s by applying the RF or RReliefF algorithm on $D^{(s)}$. In that way, each F_j is assigned an importance score W_j and a ranked list of features, R , is produced by sorting them in descending order by W_j . More specifically, let $J = [1, \dots, d]$ denote the indices of F . Then, the ranked list of features is defined as $R = [F_{j'_1}, \dots, F_{j'_d}]$ where $J' = [j'_1, \dots, j'_d], j'_j \in J$ and $W_{j'_j} \geq W_{j'_{j+1}}$. For comparison purposes, the average score of each feature F_j over all patients, i.e., \bar{W}_j and the corresponding average feature ranking R' are also calculated.

2.3.1 Random forests

RF is an ensemble of low correlated regression trees, which output is computed as the average of the individual predictions [5]. Each tree in the RF is constructed using an independent set of random vectors generated from a fixed probability distribution. Randomness is usually incorporated into the tree growing process by bootstrap resampling the original training set and randomly selecting m out of d features to split a node. The value of m is usually determined a priori equal to $d/3$.

RF provides an internal mechanism for evaluating the importance of each feature according to its contribution to the prediction of the target variable. For this purpose, the prediction error of each tree on its out-of-bag (OOB) data, i.e., the training instances that are not included in the bootstrap sample used to construct that tree, is utilized. Herein, the number of trees, B , is set to the default value of 500 provided that RF does not overfit as B increases.

The importance W_j of each feature F_j , with $j = 1, \dots, d$, is calculated as follows:

1. For each tree T_b in the RF with $b = 1, \dots, B$:
 - (a) Compute the mean squared error (MSE) of T_b on its OOB data, MSE_b .
 - (b) Permute the values of feature F_j in the OOB data of T_b and compute the new OOB error $MSE_{b,j}$.
2. The raw importance score of feature F_j is given by:

$$W_j = \frac{1}{B} \sum_{b=1}^B (MSE_{b,j} - MSE_b). \tag{1}$$

2.3.2 RReliefF

RReliefF, a classical feature ranking algorithm for regression problems, is also employed [34]. RReliefF estimates

the discriminative power of each feature F_j between adjacent instances by approximating the following difference of probabilities:

$$W_j = \frac{P_{\text{diff } C|\text{diff } F_j} P_{\text{diff } F_j} - (1 - P_{\text{diff } C|\text{diff } F_j}) P_{\text{diff } F_j}}{P_{\text{diff } C}}, \quad (2)$$

where $P_{\text{diff } C}$ corresponds to the probability two nearest instances have different predictions, $P_{\text{diff } F_j}$ corresponds to the probability that two nearest instances have different values for F_j , and $P_{\text{diff } C|\text{diff } F_j}$ corresponds to the probability that two nearest instances have different predictions and different values for F_j .

In particular, RReliefF approximates the above probabilities by iteratively (M times) selecting an instance u_m , finding its K nearest neighbors v_k and computing the following quantities:

$$Q_{\text{diff } F_j} = \sum_{m=1}^M \sum_{k=1}^K |x_j^m - x_j^k| a_{m,k}, \quad (3)$$

$$Q_{\text{diff } C} = \sum_{m=1}^M \sum_{k=1}^K |y_m - y_k| a_{m,k}, \quad (4)$$

$$Q_{\text{diff } C \& \& \text{diff } F_j} = \sum_{m=1}^M \sum_{k=1}^K |y_m - y_k| |x_j^m - x_j^k| a_{m,k}. \quad (5)$$

The city-block distance function (L_1 norm) is used to find the K nearest neighbors of u_m with respect to x_m , while K is set equal to 10. The distance between u_m and v_k is taken into account through the term $a_{m,k}$ such that closer instances have greater influence:

$$a_{m,k} = \frac{e^{-\left(\frac{\text{rank}(u_m, v_k)}{\sigma}\right)^2}}{\sum_{l=1}^K e^{-\left(\frac{\text{rank}(u_m, v_l)}{\sigma}\right)^2}}, \quad (6)$$

where $\text{rank}(u_m, v_k)$ is the position of v_k in the list of nearest neighbors of u_m sorted by distance in ascending order and σ ($\sigma = 50$ by default) is a user defined parameter. Moreover, M is set equal to its maximum value, i.e., the number of training instances.

Finally, the estimation of each W_j is given by:

$$W_j = \frac{Q_{\text{diff } C \& \& \text{diff } F_j}}{Q_{\text{diff } C}} - \frac{Q_{\text{diff } F_j} - Q_{\text{diff } C \& \& \text{diff } F_j}}{M - Q_{\text{diff } C}}. \quad (7)$$

2.4 Short-term predictive modeling of glucose concentration

Predicting glucose concentration in the s.c. space is essentially a regression problem that can be described by a linear model of the form:

$$f(x) = w^T \phi(x) + b, \quad (8)$$

in which w is a vector of parameters, ϕ is a vector of fixed nonlinear basis functions, and b is the bias parameter. The function $f: \mathfrak{R}^d \rightarrow \mathfrak{R}$ maps the input vector $x \in \mathfrak{R}^d$ to glucose concentration at time $t + l$, with t being the time at which the prediction is made and l the prediction horizon. In the present study, f is implemented through the SVR [36] and GP [31] methods, both utilizing a kernel function $k(x, x')$ rather than working directly in the transformed feature space ϕ . The parameters of the model are learned from the training set $\{(x^i, y_i)\}$, with $i = 1, \dots, N$, of each subject.

2.4.1 Support vector regression

Given a new input $x \in \mathfrak{R}^d$, the predicted by SVR glucose concentration at time $t + l$ is expressed in terms of the kernel function as follows:

$$f(x) = \sum_{i=1}^N (a_i - a_i^*) k(x, x^i) + b, \quad (9)$$

where the Lagrange multipliers a_i, a_i^* ($a_i \geq 0, a_i^* \geq 0$) are introduced in the constrained optimization process of w and, in our study, the kernel k is a Gaussian radial basis function (RBF). The sparseness of SVR solution is ensured by employing an ε -insensitive error function; the corresponding Karush–Kuhn–Tucker conditions imply that $a_i, a_i^* = 0$ for $i = 1, \dots, N$ and that all points lying inside the ε -tube have $a_i = a_i^* = 0$. Moreover, the model's complexity is controlled by the regularization parameter C which is used in the error function.

2.4.2 Gaussian processes

In the case of GP, the glucose for a new point $x \in \mathfrak{R}^d$ is estimated from a Gaussian distribution with mean and covariance given by:

$$m(x) = \sum_{i=1}^N a_i k(x, x^i), \quad \sigma^2(x) = k(x, x) + \beta^{-1} - k^T C^{-1} k, \quad (10)$$

where a_i is the i th component of $C^{-1}y$, with C denoting the $N \times N$ covariance matrix and y the target vector $y = (y_1, y_2, \dots, y_N)^T$, and the vector k has elements $k(x, x^i)$ for $i = 1, \dots, N$. The squared exponential kernel is the default one for GP regression. The noise on the observed values y is considered, and it is further assumed to be Gaussian distributed with zero mean and constant variance β for all x^i . The latter contributes to the total variance of the predictive distribution given by Eq. 10. In contrast to SVR, the kernel function k must be evaluated for all possible pairs x^i and x^j resulting in a non-sparse model.

2.5 Evaluation of feature ranking

The effectiveness of feature ranking for a specific subject s is examined with respect to the predictive performance of SVR and GP. A forward selection procedure is employed where features are sequentially added in decreasing order of importance based on RF or RReliefF ranking. To estimate the error rate of the prediction method, an external 10-fold cross-validation is applied on the dataset D_s with feature ranking following the resampling procedure itself. The latter ensures that the dataset used in the ranking process does not overlap with the test set and, therefore, reduces the selection bias in the estimates of the prediction error [2, 10, 15, 24]. The procedure used is described as follows:

1. Randomly partition $D^{(s)}$ into ten disjoint folds $D_k^{(s)}$, with $k = 1, \dots, 10$, of equal size (i.e., $N_s/10$).
2. For $k = 1, \dots, 10$:
 - (a) Let $D^{(s)} - D_k^{(s)}$ be the training set and $D_k^{(s)}$ the test set.
 - (b) Apply the RF or RReliefF algorithm to $D^{(s)} - D_k^{(s)}$ so as to produce a ranked list of features R_k .
 - (c) For $n = 1, \dots, d$:
 - i. Let $Tr_{k,n}$ and $Ts_{k,n}$ be produced from $D^{(s)} - D_k^{(s)}$ and $D_k^{(s)}$, respectively, by retaining the first n most important features according to R_k .
 - ii. Train SVR or GP glucose predictive model $M_{k,n}$ on $Tr_{k,n}$.
 - iii. Test $M_{k,n}$ on $Ts_{k,n}$ and compute the related $RMSE_{k,n}$.
3. For $n = 1, \dots, d$:
 - (a) Compute the average root-mean-squared error (RMSE) for all ten folds, i.e., $RMSE_n = \frac{1}{10} \sum_{k=1}^{10} RMSE_{k,n}$.

As the notation implies, the ranked list of features R_k can be different for each k ($k = 1, \dots, 10$). Moreover, it should be mentioned that the hyper-parameters of SVR and GP are optimized for each $Tr_{k,n}$. More specifically, the values of C , ϵ and γ minimizing the 4-fold cross-validation RMSE of SVR in $Tr_{k,n}$ are chosen by the differential evolution algorithm [38]. Regarding GP, the parameters of the squared exponential kernel along with the noise variance β are also learned for each $Tr_{k,n}$. In fact, they are internally optimized by the GP algorithm through the minimization of the negative log-likelihood function, while multiple restarts are used to alleviate the local-minimum problem.

The performance of the average feature ranking (i.e., R') is assessed in an unbiased way by precisely averaging, for each feature F_j , the scores obtained from the same fold of each 10-fold cross-validation across all patients as follows:

1. For $s = 1, \dots, 15$:
 - (a) Randomly partition $D^{(s)}$ into ten disjoint folds $D_k^{(s)}$, with $k = 1, \dots, 10$, of equal size (i.e., $N_s/10$).
2. For $k = 1, \dots, 10$:
 - (a) For $s = 1, \dots, 15$:
 - i. Compute the importance scores for all features in F by applying the RF or RReliefF algorithm to $D^{(s)} - D_k^{(s)}$. Let $W_k^{(s)} = [W_{k,1}^{(s)}, \dots, W_{k,d}^{(s)}]$ where $W_{k,j}^{(s)}$, with $j = 1, \dots, d$, is the importance score of feature F_j based on $D^{(s)} - D_k^{(s)}$.
 - (b) Compute $\bar{W}_k = [\bar{W}_{k,1}, \dots, \bar{W}_{k,d}]$ by averaging $W_{k,j}^{(s)}$, with $j = 1, \dots, d$, over s .
 - (c) Compute the ranked list R_k' by sorting \bar{W}_k in descending order.

Then, the same procedure is followed as for individualized ranking, with the difference that the dataset of each patient is not resampled and the list R_k' is used in place of R_k . In this case, the average RMSE is denoted by $RMSE_n'$.

3 Results

Figure 1 shows the average value and the standard deviation of the importance scores W_j , with $j = 1, \dots, 46$, over all 15 patients according to RF. The importance scores of features F_{1-8} and F_{9-46} are plotted separately to aid visualization. The predominance of the features corresponding to glucose concentration (i.e., F_{2-8}) is evident in both 30- and 60-min horizons, with most recent values conveying more information. The contribution of the other features to the prediction of glucose by RF is comparatively lower but not insignificant, as will be demonstrated later. In particular, their importance increases and becomes apparent for a prediction horizon of 60 min, which can be attributed to the increase of the problem complexity. Regarding gl and SEE features, which are defined with respect to the time t , their most recent values are clearly found to explain better the glucose concentration in the short term. This is the case only for Ra among the features representing the effect of meal and insulin intake (i.e., Ra , SRA and I_p) and which have been defined with respect to the time $t + l$. More specifically, it was observed that for a few patients, the values of SRA and I_p closer to the time $t + l$ are less associated with the glucose at that time.

The evaluation of features by RReliefF exhibits similar patterns as it is shown in Fig. 2. However, the difference in importance between gl and the other features is much less prominent and hh is found to discriminate adjacent samples

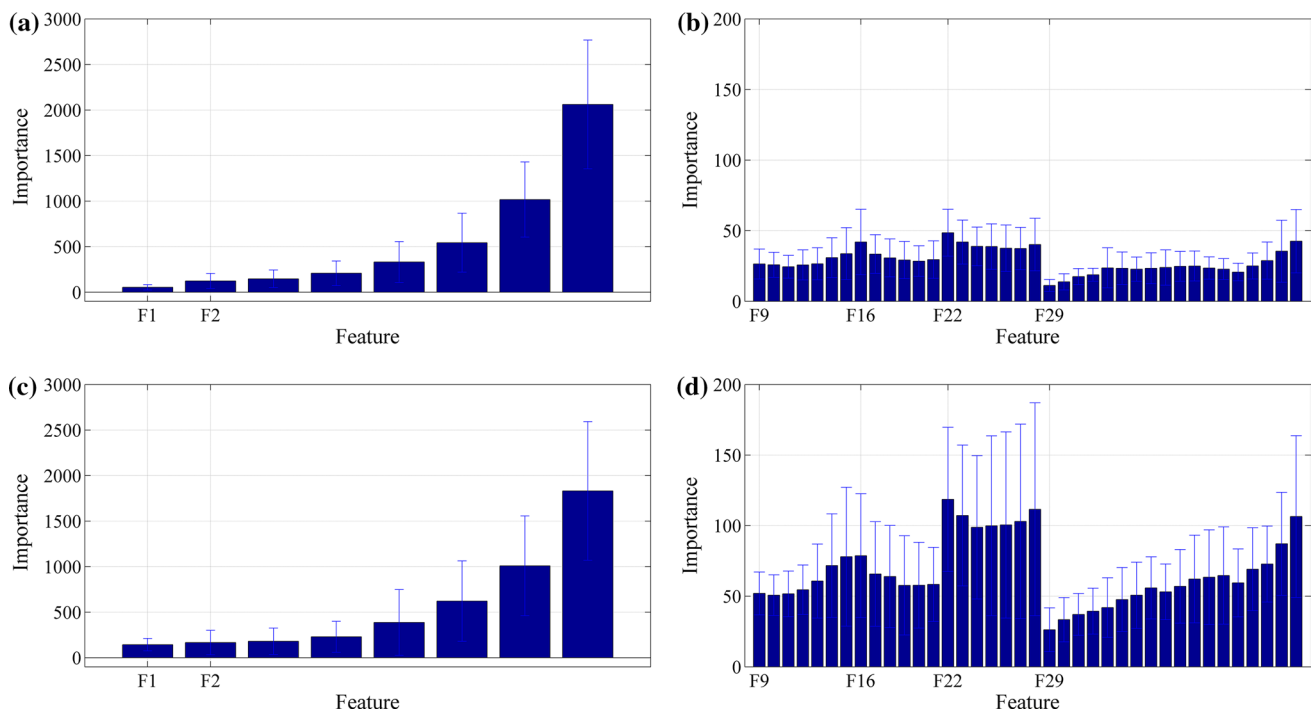


Fig. 1 Average value and standard deviation of the importance of features based on RF algorithm for (a, b) 30-min and (c, d) 60-min predictions and for all patients. The *x*-axis labels corre-

spond to the first feature of each type, i.e., $F_1 = hh$, $F_2 = gl(t - 30)$, $F_9 = Ra(t + l - 30)$, $F_{16} = SRa(t + l - 75)$, $F_{22} = I_p(t + l - 30)$ and $F_{29} = SEE(t - 170)$

equally well as gl . We also observe a smooth change in importance score over time for each type of features, and in contrast to RF, the alterations between prediction horizons are not so notable.

The image plots in Figs. 3 and 4 illustrate the ranking of features obtained by RF and RReliefF, respectively, for each individual patient. Dark shades of gray correspond to high positions in the ranking, while light shades of gray represent low positions. It is obvious that hh and the full I_p vector, in addition to gl , are ranked in the first positions for the majority of patients. On the other hand, there exist larger deviations in the ranking of the remaining features across patients, and especially of Ra and SRa . In particular for SEE , its most recent values [i.e., $SEE(t - 20)$, $SEE(t - 10)$ and $SEE(t)$] belong to the highly ranked features in more than 50 % of the patients.

In Figs. 5 and 6, the average $RMSE_n$ and $RMSE_n'$ of SVR and GP, respectively, over all 15 patients are plotted against the top-ranked ($n = 1, \dots, 46$) features for a prediction horizon of 30 min. Figures 7 and 8 illustrate the same information for 60-min horizon. We observe that the average $RMSE_n$ curve shows a sigmoidal behavior after the first iteration. Its convergence for almost $d/2$ features implies that both feature ranking algorithms properly locate high in hierarchy the most predictive features of glucose concentration. It is obvious that RReliefF outperforms RF in the

first few iterations ($n \approx 7$), which is more evident in the case of GP (with the exception of $n = 1$ where RF yields to a significantly smaller error). This could be explained by considering that RReliefF, for the majority of the patients, locates the feature hh in the first positions along with the gl values. For greater values of n , both algorithms lead to comparable average errors, although RF has systematically a slightly better performance for $n > 15$.

Similar observations hold for the $RMSE_n'$ where the average ranking of the features has been used. As it can be observed for 30-min horizon (Figs. 5 and 6), the predictive capability of the features ranked by RF in the first $n \leq 8$ positions, according to the individualized scores, is comparable with that of the average scores. After this point and until convergence is achieved, the average RF ranking, especially for SVR, yields to smaller 30-min errors compared with the individualized one. The opposite behavior is observed for RReliefF, in which the 30-min predictions with $n > 8$ best features are slightly better when the individualized scores are considered. As it is shown in Figs. 7 and 8, when the horizon increases to 60 min, the individualized RF ranking becomes superior to the average one for $n \leq 8$. This can be attributed to the fact that the average RF ranking includes the full gl vector and the hh in the first positions, whereas in the individualized case, the Ra , SRa , I_p and SEE features are occasionally included. Then, for both

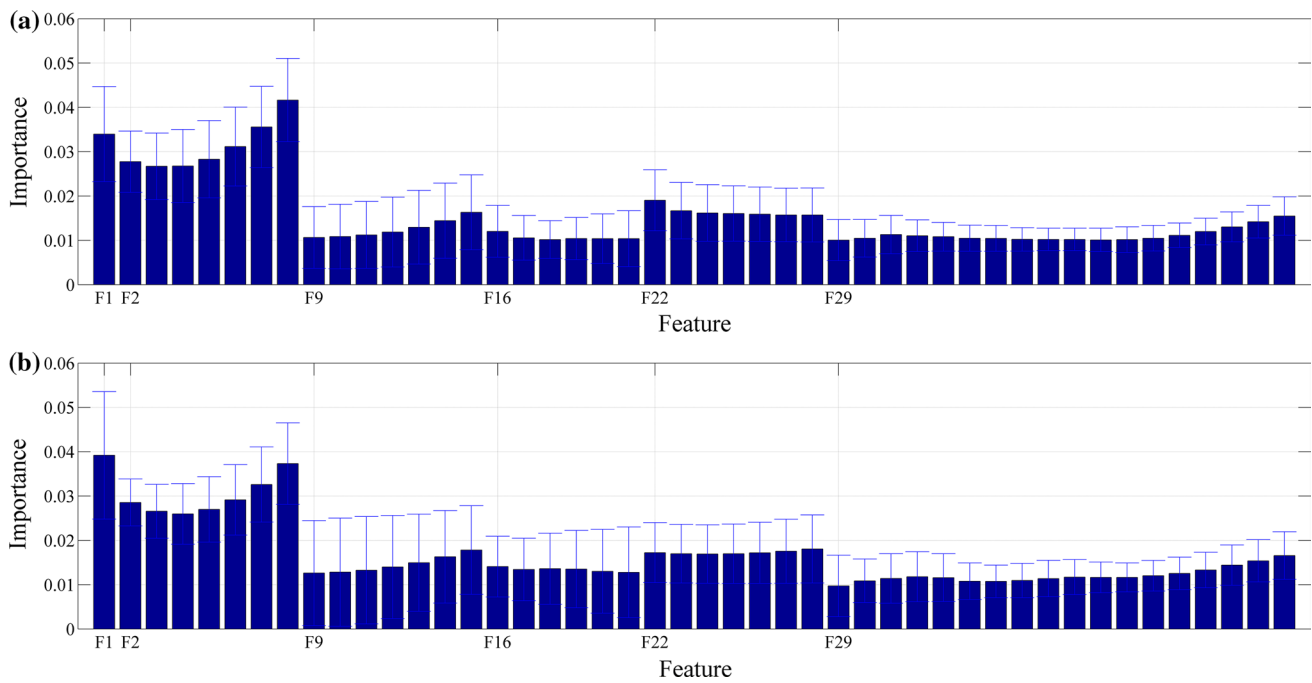


Fig. 2 Average value and standard deviation of the importance of features based on RReliefF algorithm for **a** 30-min and **b** 60-min predictions and for all patients. The *x*-axis labels correspond

to the first feature of each type, i.e., $F_1 = hh$, $F_2 = gl(t - 30)$, $F_9 = Ra(t + l - 30)$, $F_{16} = SRA(t + l - 75)$, $F_{22} = I_p(t + l - 30)$ and $F_{29} = SEE(t - 170)$

RF and RReliefF, the individualized ranking is clearly better than the average one for $n > 12$ and until convergence.

Figures 5–8 are also annotated with the average RMSE concerning three of the input cases which were defined in our previous study [18], namely case 1 (*gl*), case 4 (*hh*, *gl*, *Ra*, *SRA*, *I_p*) and case 6 (*hh*, *gl*, *Ra*, *SRA*, *I_p*, *SEE*). We can see that the average error of case 4, in which the number of features is 28, can be obtained with much less features. Moreover, a better solution can be also achieved even when the seven best features are used instead of case 1.

For each patient, the number of best features to which the average $RMSE_n$ and $RMSE'_n$ converge within 5 % of the value obtained with $n = 46$ features (i.e., n_c and n'_c , respectively) was also calculated. Table 2 presents the average value of n_c and n'_c over all patients along with the corresponding standard deviation. We can see that SVR and GP generally converge a little faster in the case of RF than in the case of RReliefF. The average values of n_c and n'_c concerning the 30-min predictions are close to each other. Nevertheless, in most of the cases, the 60-min error curves converge considerably faster when features are ranked individually for each subject.

4 Discussion

In this paper, a study on the evaluation of short-term predictors of s.c. glucose concentration in type 1 diabetes was

presented. This problem was addressed for the first time in the literature with the aid of RF and RReliefF algorithms, which were applied to self-monitoring data. Their efficacy was verified with respect to the predictive performance of two machine-learning regression models.

The need to augment the input of predictive models with features able to reveal the daily dynamics of glucose concentration is critical. The utilization of information on meals, insulin therapy and physical activities, besides glucose time series, has been shown to lower the prediction error [8, 13, 18, 41, 43, 46–48]. The proposed dataset, in addition to the glucose signal and the time of the day, includes some novel features highly connected to glucose dynamics. First, the future values of the *Ra* and *I_p* simulated signals were used by expanding the simulation time from the present time up to the time for which the prediction is to be made and provided that that no future event (i.e., meal, insulin injection) will occur during that period. This approach was also followed in [47] for the *Ra* signal with the difference that meal information should be announced by the patient *l* min in advance. Nevertheless, the area under the *Ra* curve has not been introduced elsewhere as a predictor variable. Similarly, the variable *SEE*, which represents the cumulative energy expenditure over time, was first introduced in [18]. Actually, the few studies using information from a physical activity monitor for making predictions are based only on the past instantaneous values

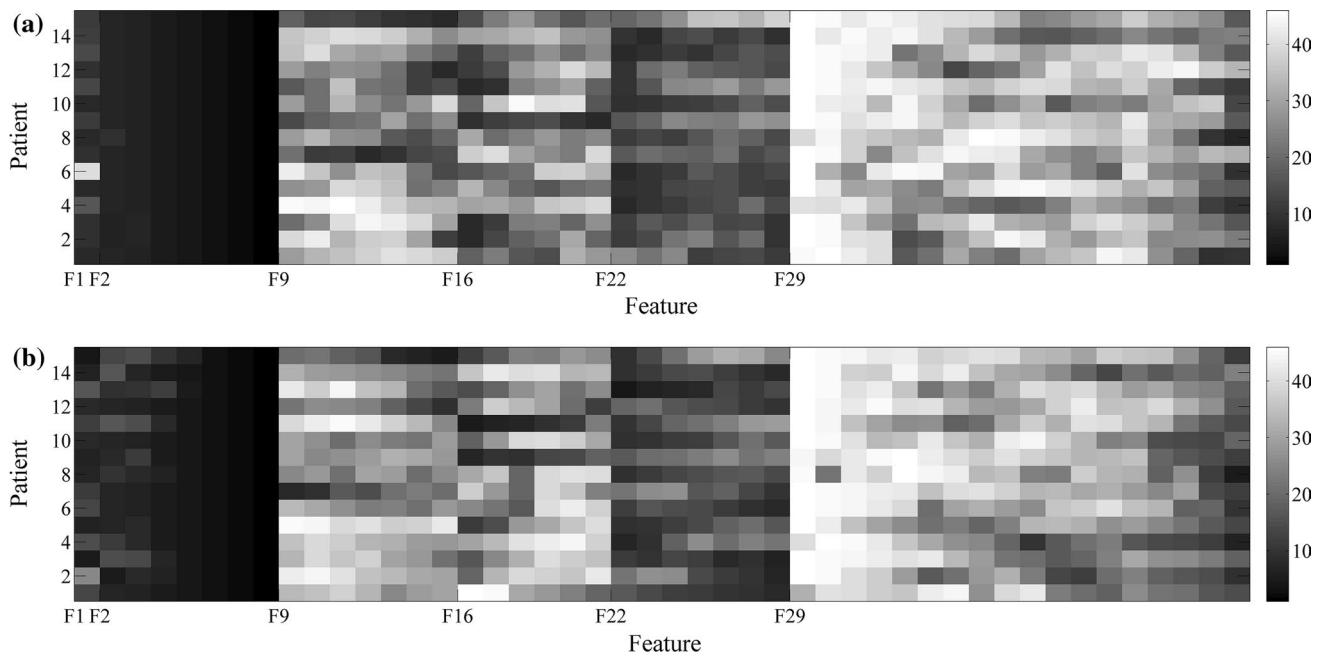


Fig. 3 Image plot of RF ranking for each patient for **a** 30-min and **b** 60-min prediction horizon. Darker shades of gray indicate a higher-ranking position, and lighter shades of gray repre-

sent a lower one. The x -axis labels correspond to the first feature of each type, i.e., $F_1 = hh$, $F_2 = gl(t - 30)$, $F_9 = Ra(t + l - 30)$, $F_{16} = SRa(t + l - 75)$, $F_{22} = I_p(t + l - 30)$ and $F_{29} = SEE(t - 170)$

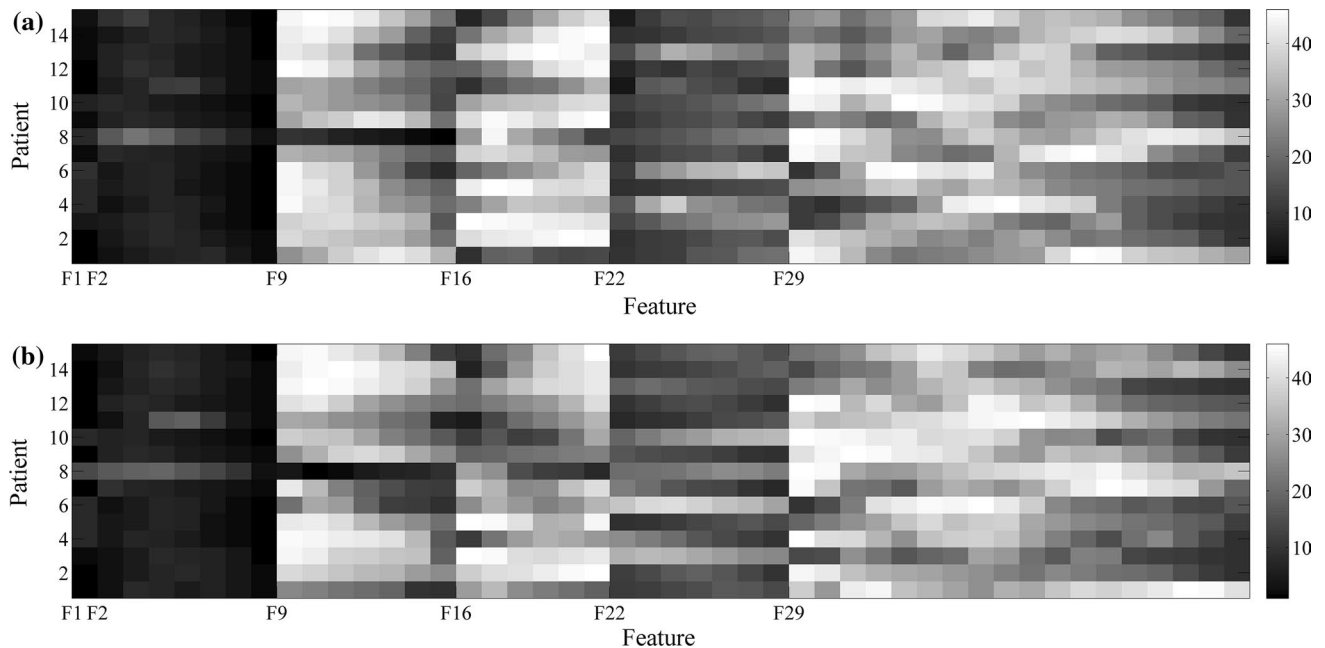


Fig. 4 Image plot of RReliefF ranking for each patient for **a** 30-min and **b** 60-min prediction horizon. Darker shades of gray indicate a higher-ranking position, and lighter shades of gray repre-

sent a lower one. The x -axis labels correspond to the first feature of each type, i.e., $F_1 = hh$, $F_2 = gl(t - 30)$, $F_9 = Ra(t + l - 30)$, $F_{16} = SRa(t + l - 75)$, $F_{22} = I_p(t + l - 30)$ and $F_{29} = SEE(t - 170)$

of physiological signals (e.g., energy expenditure, galvanic skin response, heat flux) [13, 41]. However, the effect of all these variables on glucose metabolism varies considerably

among type 1 diabetes patients due to a combination of environmental and biological factors. In addition, the efficient representation of the temporal dependencies between

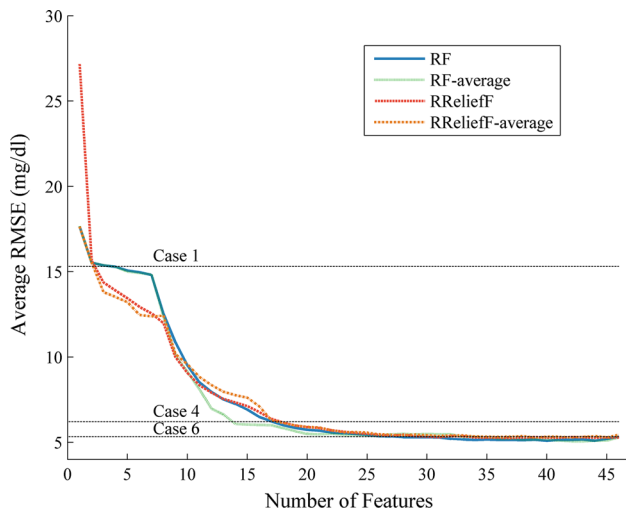


Fig. 5 Average 10-fold cross-validation RMSE rate of SVR regression models over all 15 patients against the top-ranked features identified by RF and RReliefF for prediction horizon of 30 min

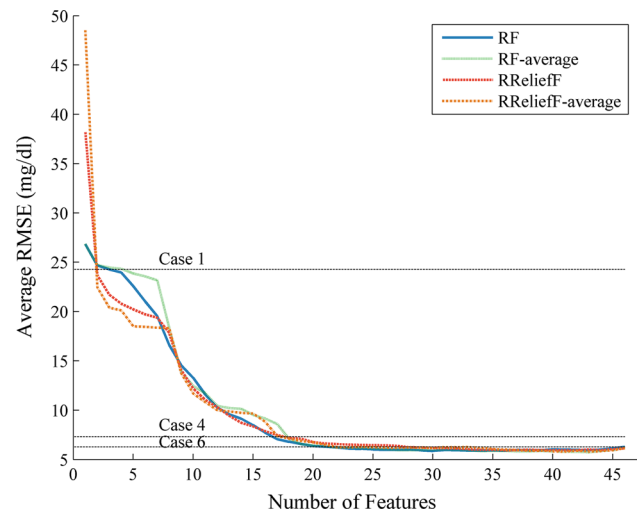


Fig. 7 Average 10-fold cross-validation RMSE rate of SVR regression models over all 15 patients against the top-ranked features identified by RF and RReliefF for prediction horizon of 60 min

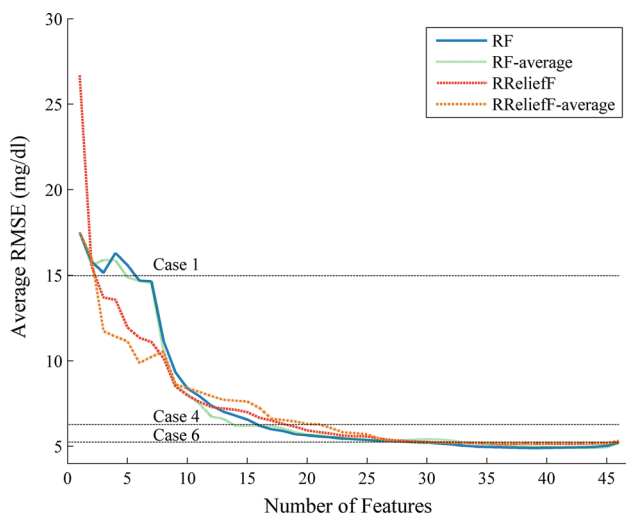


Fig. 6 Average 10-fold cross-validation RMSE rate of GP regression models over all 15 patients against the top-ranked features identified by RF and RReliefF for prediction horizon of 30 min

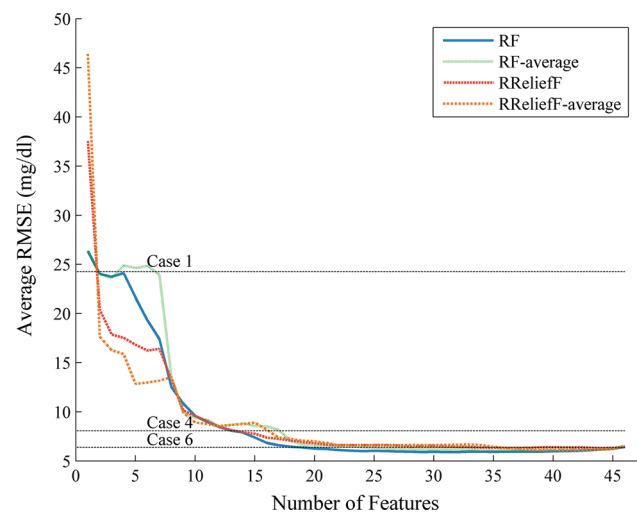


Fig. 8 Average 10-fold cross-validation RMSE rate of GP regression models over all 15 patients against the top-ranked features identified by RF and RReliefF for prediction horizon of 60 min

the input variables and the glucose concentration can be challenging. On this basis, we attempted to evaluate separately for each patient the proposed feature set, whose predictive capacity has already been validated as a whole. This is the main novelty of this work; since to the authors' knowledge, there has been no other attempt to determine and assess the importance of such a multivariate feature set for predicting glucose at the individual level.

RF and RReliefF are two entirely different feature ranking algorithms but both are well suited for regression problems [34, 39]. The importance score computed for each

feature by RF expresses the increase in the OOB prediction error when its values are randomly permuted (to mimic its absence in the prediction). It should be mentioned that the OOB error is an unbiased estimation of the generalization error of RF, which converges as the number of trees increases [5]. On the other hand, RReliefF is a statistical approach that approximates the probabilities of (non) separation of near instances by a given feature across the problem space. An appealing property of both algorithms is that they are context sensitive, i.e., they take into account all attributes when estimating their importance. More

Table 2 Number of features to which the RMSE rate converges for all 15 subjects

	Prediction horizon			
	30 min		60 min	
	SVR	GP	SVR	GP
Individualized ranking				
RF	22.3 ± 7.0	22.5 ± 7.7	21.5 ± 6.8	18.1 ± 8.9
RReliefF	24.5 ± 8.0	24.8 ± 9.2	24.7 ± 10.2	22.3 ± 10.6
Average ranking				
RF	21.4 ± 7.5	22.9 ± 9.7	26.5 ± 9.5	24.6 ± 12.6
RReliefF	24.9 ± 5.9	25.6 ± 7.4	23.5 ± 7.6	26.4 ± 11.6

Data are mean ± standard deviation values

specifically, RF can efficiently learn the relationships hidden in the dataset, while RReliefF detects existing dependencies in the feature space by exploiting the distance between instances. As a result, both algorithms behave well in the presence of groups of highly correlated features, which is the case in our problem. Moreover, as opposed to shrinkage methods for linear regression [32], RF and RRelief do not assume a linear and sparse (with many zero regression coefficients) model. This is of particular importance in glucose predictive modeling where linear and nonlinear components of glucose dynamics should be described. Another important class of embedded methods use the change in the objective function when one feature is removed or added as a ranking criterion, and in combination with a greedy search strategy, they yield nested subsets of features, e.g., recursive feature elimination [20]. However, in this approach, the prediction technique should be retrained for each new subset of features (i.e., d times in stepwise feature selection), whereas RF needs to be fitted to the training set only once. Note that the computational complexity of RF (i.e., $O(B \cdot m \cdot N \cdot \log N)$ with $m = d/3$) can be considered comparable to that of RReliefF (i.e., $O(d \cdot N \cdot \log N)$), despite being an embedded method.

The way the two feature ranking algorithms operate is definitely reflected in their output. In particular, RF's output reveals (1) the infeasibility of predicting the s.c. glucose concentration without exploiting its recent values (e.g., the average OOB MSE increases by ≈ 2000 mg/dl when $gl(t)$ is randomly permuted) and (2) the more pronounced effect of the other features with increasing prediction horizon. On the other hand, the fact that RReliefF computes the discriminative (and not the predictive) ability of each feature being; however, aware of the context of other features can explain (1) the lack of great differences in features' scores, (2) the similarity of the output between prediction horizons and (3) the smooth transition in scores over time for features of the same type.

Both RF and RRelief highlighted how essential is the s.c. glucose signal itself for both prediction horizons and for all 15 patients. Of great interest is that the time at which the predictions are made (i.e., hh) is systematically located in the first positions and its score is comparable to that of glucose. This reflects the existence of daily (24-h) patterns in glucose time series which are imposed either by each patient's lifestyle or by circadian rhythms related to glucose homeostasis [14, 21]. The contribution of the other features was also well demonstrated, with I_p features outweighing on average Ra , SRA and SEE ones. In addition, both algorithms, and especially RF, revealed some rational attenuation trends over time in average scores of gl , Ra and SEE [9, 23, 25, 35]. The effect of I_p seems to be less immediate (since its scores tend to decrease as getting closer to the time of prediction $t + l$), which can be considered consistent with clinical evidence indicating inherent delays in peripheral and hepatic insulin action [26]. Moreover, the results support the existence of substantial inter-patient differences.

Short-term predictive modeling of the s.c. glucose concentration using SVR and GP further verified the quality of the resulting feature ranking. The behavior of the average $RMSE_n$ curve did confirm that the top-ranked features constitute the best predictors of glucose in the examined feature set. The fact that the prediction performance did not degenerate by applying the average feature ranking reveals the generalizability and robustness of the results. In particular, individualized feature ranking was found to be more appropriate for 60-min predictions, which may suggest personalized glucose predictive approaches are preferable as prediction horizon increases. Regarding the short-term glucose dynamics (i.e., 30-min horizon), the convergence rate of the average $RMSE_n'$ error curve was similar with that of the average $RMSE_n$ curve. Moreover, the convergence of both error curves for a considerably smaller than $d = 46$ number of features, and the consequent reduction of the input size, is indeed of paramount importance for regression analysis. Nevertheless, we did not find a certain point after which the average error starts to increase, which is mainly due to the fact that all features are relevant to the studied problem. We should mention at this point that the two kernel-based techniques were chosen due to their high prediction accuracy in both normal and critical glucose value regions [19]. However, other well-established machine-learning regression techniques could be also applied.

Table 3 presents a comparison of the proposed work with other literature studies utilizing a multivariate dataset. Note that the prediction error provided for SVR and GP corresponds to the first $n = 20$ best features. A direct comparison of the presented results is not fair since they have been derived by different training/testing approaches. In [46, 47], the predictive models are tested on patients not

Table 3 Comparison with other methods reported in the literature

Study	Method	Feature set (number of features)	Dataset	Prediction horizon (min)/RMSE (mg/dl)
Zecchin et al. [47]	Feed-forward neural network and first-order polynomial model	CGM data, glucose rate of appearance after a meal (8)	15 Type 1 diabetic subjects	30/14.0 ± 4.1
Eren-Oruklu et al. [13]	Recursive ARMAX model	CGM data, energy expenditure, average longitudinal acceleration, near-body temperature, heat flux, galvanic skin response (15)	5 Type 2 diabetic subjects	30/4.2 ± 5.1 ^a
Turksoy et al. [41]	Recursive ARMAX model	CGM data, insulin on board, energy expenditure, galvanic skin response (20)	14 Type 1 diabetic subjects	30/11.7 60/34.7
Daskalaki et al. [8]	Online adaptive RNN	CGM data and insulin pump infusion rate (N/A, patient-specific)	23 Type 1 diabetic subjects	15/11.9 30/18.9 45/26.1
Zecchin et al. [46]	Jump neural network model	CGM data, glucose rate of appearance after a meal (4)	20 Type 1 diabetic subjects	30/16.6 ± 3.1
This work	SVR—RF	CGM data, glucose rate of appearance after a meal, plasma insulin concentration, energy expenditure, time of the day (20)	15 Type 1 diabetic subjects	30/5.7 ± 1.5 60/6.4 ± 2.1
	SVR—RRF			30/5.9 ± 1.4 60/6.8 ± 2.0
	GP—RF			30/5.6 ± 1.7 60/6.3 ± 2.6
	GP—RRF			30/5.9 ± 1.6 60/6.8 ± 2.9

Data are mean ± standard deviation values

^a This value refers to relative absolute deviation RAD(%)

included in the training set, while those in [8, 13, 41] are recursively trained on each patient dataset. Our results indicate that multivariate nonlinear regression models can provide predictions of high accuracy, which is also in agreement with previous findings. Moreover, we can see that the inclusion of information on physical activities is able to improve performance even when a linear model is adopted. The main difference of our work is that the input is not pre-defined, but it is selected separately for each patient from a high-dimensional feature set which may result in much simpler models.

The fact that RF and RReliefF algorithms yield consistent results across multiple subjects for both 30- and 60-min prediction horizons implies their potential for use as an exploratory tool in the predictive analysis of type 1 diabetes data. Given the monitoring data of a new unseen patient, these algorithms can be applied to obtain a first reliable estimate of the predictive capability of the input variables. Certainly, the low computational complexity of feature ranking allows one to investigate longer latency time intervals than those examined in this study as well as to examine the impact of new descriptive features. The specification of the dimension of the input with respect to a regression technique requires employing the forward selection procedure, not necessarily in an exhaustive way, but for some subsets of features until the error converges. Similarly, the average ranking of the features could be utilized in the construction of “generalized” predictive models from the entire patients’ set. Again, the precise merging of the same folds of each 10-fold cross-validation across all patients would be needed to ensure unbiased estimates of the prediction error. As a future work, RF and RReliefF need to be evaluated in a large number of patients over a long period of time. To this end, both algorithms could be also tested on patients who are monitored during different time periods to investigate how consistent are the results for a patient and what is the effect of lifestyle or physiological changes. In any case, the clinicians should interpret the calculated set of best features together with other clinical information.

5 Conclusions

In this work, we proposed the application of RF and RReliefF feature evaluation algorithms on real-life type 1 diabetes data as a means to customize the input of glucose predictive models. Both algorithms produced rational, robust results revealing not only the global importance of features concerning s.c. glucose profile, time of prediction and plasma insulin concentration but also the different role of food intake and physical activity among patients. A very interesting finding deserving more attention in the future

was that the plasma insulin concentration is systematically found to outweigh the rate of appearance as well as the cumulative amount of meal-derived glucose inserted in the plasma over time. In addition, the possibility of obtaining equally accurate predictions using on average less than half of the initial features was demonstrated by utilizing the derived feature ranking in the development of SVR and GP predictive models. It was shown that RF and RReliefF result in equally predictive feature ranking, but our foremost conclusion is that both show a consistent behavior across all patients.

Acknowledgments This work is supported by the research project “Development of an Information Environment for Diabetes Data Analysis and New Knowledge Mining” that has been co-financed by the European Union [European Regional Development Fund (ERDF)] and Greek national funds through the Operational Program “THESALY-MAINLAND GREECE AND EPIRUS-2007-2013” of the National Strategic Reference Framework (NSRF 2007–2013).

References

1. Abu-Rmieleh A, Garcia-Gabin W, Zambrano D (2010) A robust sliding mode controller with internal model for closed-loop artificial pancreas. *Med Biol Eng Comput* 48(12):1191–1201. doi:[10.1007/s11517-010-0665-3](https://doi.org/10.1007/s11517-010-0665-3)
2. Ambroise C, McLachlan GJ (2002) Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Nat Acad Sci USA* 99(10):6562–6566. doi:[10.1073/pnas.102102699](https://doi.org/10.1073/pnas.102102699)
3. American Diabetes A (2014) Standards of medical care in diabetes-2014. *Diabetes Care* 37(Suppl 1):S14–80. doi:[10.2337/dc14-S014](https://doi.org/10.2337/dc14-S014)
4. Bergenstal RM, Klonoff DC, Garg SK, Bode BW, Meredith M, Slover RH, Ahmann AJ, Welsh JB, Lee SW, Kaufman FR, Group AI-HS (2013) Threshold-based insulin-pump interruption for reduction of hypoglycemia. *New Engl J Med* 369(3):224–232. doi:[10.1056/NEJMoa1303576](https://doi.org/10.1056/NEJMoa1303576)
5. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. doi:[10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)
6. Buckingham B, Cameron F, Calhoun P, Maahs DM, Wilson DM, Chase HP, Bequette BW, Lum J, Sibayan J, Beck RW, Kollman C (2013) Outpatient safety assessment of an in-home predictive low-glucose suspend system with type 1 diabetes subjects at elevated risk of nocturnal hypoglycemia. *Diabetes Technol Ther* 15(8):622–627. doi:[10.1089/dia.2013.0040](https://doi.org/10.1089/dia.2013.0040)
7. Cryer PE (2008) The barrier of hypoglycemia in diabetes. *Diabetes* 57(12):3169–3176. doi:[10.2337/db08-1084](https://doi.org/10.2337/db08-1084)
8. Daskalaki E, Norgaard K, Zuger T, Prountzou A, Diem P, Mouggiakakou S (2013) An early warning system for hypoglycemic/hyperglycemic events based on fusion of adaptive prediction models. *J Diabetes Sci Technol* 7(3):689–698
9. Derouich M, Boutayeb A (2002) The effect of physical exercise on the dynamics of glucose and insulin. *J Biomech* 35(7):911–917
10. Di Camillo B, Sanavia T, Martini M, Jurman G, Sambo F, Barla A, Squillario M, Furlanello C, Toffolo G, Cobelli C (2012) Effect of size and heterogeneity of samples on biomarker discovery: synthetic and real data assessment. *PLoS ONE*. doi:[10.1371/journal.pone.0032200](https://doi.org/10.1371/journal.pone.0032200)
11. Dua P, Doyle FJ 3rd, Pistikopoulos EN (2009) Multi-objective blood glucose control for type 1 diabetes. *Med Biol Eng Comput* 47(3):343–352. doi:[10.1007/s11517-009-0453-0](https://doi.org/10.1007/s11517-009-0453-0)

12. Eren-Oruklu M, Cinar A, Quinn L (2010) Hypoglycemia prediction with subject-specific recursive time-series models. *J Diabetes Sci Technol* 4(1):25–33
13. Eren-Oruklu M, Cinar A, Rollins DK, Quinn L (2012) Adaptive System Identification for estimating future glucose concentrations and hypoglycemia alarms. *Automatica: J IFAC* 48(8):1892–1897. doi:10.1016/j.automatica.2012.05.076
14. Froy O (2010) Metabolism and circadian rhythms—implications for obesity. *Endocr Rev* 31(1):1–24. doi:10.1210/er.2009-0014
15. Furlanello C, Serafini M, Merler S, Jurman G (2003) Entropy-based gene ranking without selection bias for the predictive classification of microarray data. *BMC Bioinform.* doi:10.1186/1471-2105-4-54
16. Gani A, Gribok AV, Rajaraman S, Ward WK, Reifman J (2009) Predicting subcutaneous glucose concentration in humans: data-driven glucose modeling. *IEEE Trans Bio-Med Eng* 56(2):246–254. doi:10.1109/TBME.2008.2005937
17. Georga E, Protopappas V, Guillen A, Fico G, Ardigo D, Arredondo MT, Exarchos TP, Polyzos D, Fotiadis DI (2009) Data mining for blood glucose prediction and knowledge discovery in diabetic patients: the METABO diabetes modeling and management system. In: Conference proceedings: annual international conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference 2009, pp 5633–5636. doi:10.1109/IEMBS.2009.5333635
18. Georga E, Protopappas VC, Ardigo D, Marina M, Zavaroni I, Polyzos D, Fotiadis D (2012) Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression. *IEEE J Biomed Health Inform.* doi:10.1109/TITB.2012.2219876
19. Georga EI, Protopappas VC, Ardigo D, Polyzos D, Fotiadis DI (2013) A glucose model based on support vector regression for the prediction of hypoglycemic events under free-living conditions. *Diabetes Technol Ther* 15(8):634–643. doi:10.1089/dia.2012.0285
20. Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. *Mach Learn* 46(1–3):389–422. doi:10.1023/A:1012487302797
21. Huang W, Ramsey KM, Marcheava B, Bass J (2011) Circadian rhythms, sleep, and metabolism. *J Clin Investig* 121(6):2133–2141. doi:10.1172/JCI146043
22. Hughes CS, Patek SD, Breton MD, Kovatchev BP (2010) Hypoglycemia prevention via pump attenuation and red–yellow–green “traffic” lights using continuous glucose monitoring and insulin pump data. *J Diabetes Sci Technol* 4(5):1146–1155
23. Kovatchev B, Clarke W (2008) Peculiarities of the continuous glucose monitoring data stream and their impact on developing closed-loop control technology. *J Diabetes Sci Technol* 2(1):158–163
24. Krzanowski WJ (1993) Discriminant-analysis and statistical pattern-recognition—Mclachlan, GJ. *J Classif* 10(1):128–130
25. Lehmann ED, Deutsch T (1992) A physiological model of glucose-insulin interaction in type 1 diabetes mellitus. *J Biomed Eng* 14(3):235–242
26. Miles PD, Levisetti M, Reichart D, Khoursheed M, Moossa AR, Olefsky JM (1995) Kinetics of insulin action in vivo. Identification of rate-limiting steps. *Diabetes* 44(8):947–953
27. Mougiakakou SG, Bartsocas CS, Bozas E, Chaniotakis N, Iliopoulou D, Kouris I, Pavlopoulos S, Prountzou A, Skevofilakas M, Tsoukalis A, Varotsis K, Vazeou A, Zarkogianni K, Nikita KS (2010) SMARTDIAB: a communication and information technology approach for the intelligent monitoring, management and follow-up of type 1 diabetes patients. *IEEE Trans Inform Technol Biomed: Publ IEEE Eng Med Biol Soc* 14(3):622–633. doi:10.1109/TITB.2009.2039711
28. Perez-Gandia C, Facchinetti A, Sparacino G, Cobelli C, Gomez EJ, Rigla M, de Leiva A, Hernando ME (2010) Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring. *Diabetes Technol Ther* 12(1):81–88. doi:10.1089/dia.2009.0076
29. Phillip M, Battelino T, Atlas E, Kordonouri O, Bratina N, Miller S, Biester T, Avbelj Stefanija M, Muller I, Nimri R, Danne T (2013) Nocturnal glucose control with an artificial pancreas at a diabetes camp. *N Engl J Med* 368(9):824–833. doi:10.1056/NEJMoa1206881
30. Pickup JC (2012) Insulin-pump therapy for type 1 diabetes mellitus. *New Engl J Med* 366(17):1616–1624. doi:10.1056/NEJMct1113948
31. Rasmussen CE, Nickisch H (2010) Gaussian processes for machine learning (GPML) toolbox. *J Mach Learn Res* 11:3011–3015
32. Rasmussen MA, Bro R (2012) A tutorial on the Lasso approach to sparse modeling. *Chemometr Intell Lab* 119:21–31. doi:10.1016/j.chemolab.2012.10.003
33. Reifman J, Rajaraman S, Gribok A, Ward WK (2007) Predictive monitoring for improved management of glucose levels. *J Diabetes Sci Technol* 1(4):478–486
34. Robnik-Sikonja M, Kononenko I (2003) Theoretical and empirical analysis of ReliefF and RReliefF. *Mach Learn* 53(1–2):23–69. doi:10.1023/A:1025667309714
35. Roy A, Parker RS (2007) Dynamic modeling of exercise effects on plasma glucose and insulin levels. *J Diabetes Sci Technol* 1(3):338–347
36. Smola AJ, Scholkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14(3):199–222. doi:10.1023/B:Stco.0000035301.49549.88
37. Sparacino G, Zanderigo F, Corazza S, Maran A, Facchinetti A, Cobelli C (2007) Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. *IEEE T Bio-Med Eng* 54(5):931–937. doi:10.1109/Tbme.2006.889774
38. Storn R, Price K (1997) Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim* 11(4):341–359. doi:10.1023/A:1008202821328
39. Strobl C, Malley J, Tutz G (2009) An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods* 14(4):323–348. doi:10.1037/a0016973
40. Tarin C, Teufel E, Pico J, Bondia J, Pfeiderer HJ (2005) Comprehensive pharmacokinetic model of insulin Glargine and other insulin formulations. *IEEE Trans Bio-Med Eng* 52(12):1994–2005. doi:10.1109/TBME.2005.857681
41. Turksoy K, Bayrak ES, Quinn L, Littlejohn E, Rollins D, Cinar A (2013) Hypoglycemia early alarm systems based on multivariable models. *Ind Eng Chem Res.* doi:10.1021/ie3034015
42. Turksoy K, Quinn L, Littlejohn E, Cinar A (2014) Multivariable adaptive identification and control for artificial pancreas systems. *IEEE Trans Bio-Med Eng* 61(3):883–891. doi:10.1109/TBME.2013.2291777
43. Yamaguchi M, Kaseda C, Yamazaki K, Kobayashi M (2006) Prediction of blood glucose level of type 1 diabetics using response surface methodology and data mining. *Med Biol Eng Comput* 44(6):451–457. doi:10.1007/s11517-006-0049-x
44. Zarkogianni K, Vazeou A, Mougiakakou SG, Prountzou A, Nikita KS (2011) An insulin infusion advisory system based on auto-tuning nonlinear model-predictive control. *IEEE Trans Bio-Med Eng* 58(9):2467–2477. doi:10.1109/TBME.2011.2157823
45. Zecchin C, Facchinetti A, Sparacino G, Cobelli C (2013) Reduction of number and duration of hypoglycemic events by glucose prediction methods: a proof-of-concept in silico study. *Diabetes Technol Ther* 15(1):66–77. doi:10.1089/dia.2012.0208

46. Zecchin C, Facchinetti A, Sparacino G, Cobelli C (2014) Jump neural network for online short-time prediction of blood glucose from continuous monitoring sensors and meal information. *Comput Methods Programs Biomed* 113(1):144–152. doi:[10.1016/j.cmpb.2013.09.016](https://doi.org/10.1016/j.cmpb.2013.09.016)
47. Zecchin C, Facchinetti A, Sparacino G, De Nicolao G, Cobelli C (2012) Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration. *IEEE Trans Bio-Med Eng* 59(6):1550–1560. doi:[10.1109/TBME.2012.2188893](https://doi.org/10.1109/TBME.2012.2188893)
48. Zhao C, Dassau E, Jovanovic L, Zisser HC, Doyle FJ 3rd, Seborg DE (2012) Predicting subcutaneous glucose concentration using a latent-variable-based statistical method for type 1 diabetes mellitus. *J Diabetes Sci Technol* 6(3):617–633