2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

An Ensemble Learning Algorithm Based on Dynamic Voting for Targeting the Optimal Insulin Dosage in Type 1 Diabetes Management*

Giulia Noaro, Giacomo Cappon, Giovanni Sparacino, Andrea Facchinetti

Abstract—**People with type 1 diabetes (T1D) need exogenous insulin administrations several times a day. The amount of injected insulin is key for maintaining the concentration of blood glucose (BG) within a physiological safe range. According to well-established clinical guidelines, insulin dosing at mealtime is calculated through an empirical formula which, however, does not take advantage of the knowledge of BG trend provided in real-time by continuous glucose monitoring (CGM) sensors. To overcome suboptimal insulin dosage, we recently used machine learning techniques to build two new models, one linear and one nonlinear, which incorporate BG trend information.**

In this work, we propose an ensemble learning method for mealtime insulin bolus estimation based on dynamic voting, which combines the two models by taking advantage of where each alternative performs better. Being the resulting model black-box, a tool that enables its interpretability was applied to evaluate the contribution of each feature. The proposed model was trained using a synthetic dataset having information on 100 virtual subjects with different mealtime conditions, and its performance was evaluated within a simulated environment.

The benefit given by the ensemble method compared to the single models was confirmed by the high time within the target glycemic range, and the trade-off reached in terms of time spent below and above this range. Moreover, the model interpretation pointed out the key role played by the information on BG dynamics in the estimation of insulin dosage.

Clinical Relevance— **The proposed approach provides an effective and safe rule for the computation of mealtime insulin dosage in T1D management.**

I. Introduction

Type 1 diabetes (T1D) is an autoimmune disease which consists of pancreatic beta-cells not producing insulin [1], thus resulting in high blood glucose (BG) concentration. As a consequence, people with T1D need to administer multiple boluses of insulin exogenously to counteract this condition. On the other hand, overdosage of exogenous insulin leads to hypoglycemia (BG < 70 mg/dL) [2]. As such, to maintain the BG level within the target range (70 ≤ BG ≤ 180 mg/dL), the precise estimation of the insulin bolus (IB) is crucial. Typically, this quantity is calculated through a standard formula (SF) [3]:

$$IB_{SF} = \frac{CHO}{CR} + \frac{Gc-Gt}{CF} - IOB \quad (1)$$

where IB_{SF} (U) is the total IB amount, CHO (g) is the meal carbohydrates intake, CR (g/U) is the insulin-to-carbohydrates ratio [4], G_c (mg/dL) is the current BG concentration, usually obtained by self-monitoring fingerprick devices, G_t (mg/dL) is the target BG, CF (mg/dL/U) is the correction factor [4], and IOB (U) is the so-called insulin-on-board, i.e., an estimate of the amount of insulin still acting on the body from previous administrations [5].

Though well established in clinical practices, the formula of Eq.(1) does not take into account glucose dynamics, e.g., the BG derivative ($\dot{BG}$), an information accessible through the use of modern devices, the so-called minimally-invasive continuous glucose monitoring (CGM) sensors, which are revolutionizing T1D management [6], [7]. Intuitively, if $\dot{BG}$ at mealtime is positive/negative the IB dosing should be increased/decreased accordingly. Hence, the aim being optimizing IB calculation by accounting for $\dot{BG}$, we recently proposed two new modeling approaches [8], [9], one linear and one nonlinear, that proved to outperform the SF in silico. However, a further margin of improvement may come from merging these two techniques in a single one, thus developing an ensemble model. Indeed, the different outputs can be combined by applying a specific model within the "local region" where the regression error is the lowest, thus taking advantage that each model produces better results inside certain subareas of the application domain.

Following this rationale, in this work, we merged [8], [9] to build an ensemble learning method based on dynamic voting (DV) [10], i.e., the final prediction is derived by weighting the two outputs according to the local performance, with the aim of developing a more effective model for the estimation of the optimal IB dose. The proposed methodology led to satisfactory results in silico, suggesting that a further investigation in this direction is needed. In addition, the model interpretation using SHapley Additive exPlanations (SHAP), i.e., a novel tool to interpret model predictions, confirmed the importance of $\dot{BG}$ within the model, proving that the BG dynamic is a valuable information for IB optimization.

II. Dataset

A. Simulated dataset generation

Learning and assessment of the methodology were performed by generating synthetic data of 100 virtual adult

*Part of this work was supported by MIUR (Italian Minister for Education) under the initiative "Department of Excellence" (Law 232/2016).

G.Noaro, G. Cappon, G.Sparacino, A. Facchinetti are with the Department of Information Engineering, University of Padova, G. Gradenigo 6/B, 35131, Padova, Italy. (phone +39 049 827 7669; email: noarogiuli, cappongi, gianni, facchine@dei.unipd.it).

subjects using the UVa/Padova T1D Simulator [11]. The virtual population underwent multiple single-meal scenarios in a noise-free environment, i.e., error sources for the CHO counting, BG measurement, and $\dot{BG}$ estimation were not taken into account. Moreover, no corrective actions such as CHO intakes or correction IB were allowed during the postprandial window. This experimental set-up was chosen to avoid any confounders, which could have influenced our study.

Each single-meal simulation had a duration of 12 hours, in which the first 6 hours were used to obtain the specific mealtime conditions in terms of BG and $\dot{BG}$. Particularly, we brought each virtual subject to prandial BG levels taking values between 70 and 180 mg/dL with a step of 10 mg/dL, and $\dot{BG}$ from -2 to +2 mg/dL/min with a step of 0.5 mg/dL/min. Then, we set for each subject and prandial condition a different amount of meals, ranging between 10 and 150 grams of CHO, with step of 10 g. Finally, we computed the optimal IB (IB_{OPT}), i.e., the target of our learning algorithm, by minimizing the blood glucose risk index (BGRI) [12] within the 6 hours postprandial window.

After this procedure, we extracted 10 features related both to the mealtime condition (G_c, CHO, $\dot{BG}$, IOB, SF, basal insulin (I_b)) and the physiology of the virtual subject (CR, CF, G_t, body weight (BW)). Then, the whole dataset was divided into a training and test set with a 4:1 ratio, in a way that the records related to the same virtual subject belongs or to the training or to the test set, to provide an unbiased evaluation. Moreover, the variables were preprocessed by removing the mean and scaling to unit variance.

III. Methods

A. Models composing the ensemble

We selected as base regressors, i.e., as models that composed the ensemble, the least absolute shrinkage and selection operator (LASSO) model [8] and the random forest (RF) model [9]. The LASSO model allows both variable selection and regularization, by adding a penalty term to the loss function. On the other hand, the RF model, i.e., a learning method based on an ensemble of decision trees, is focused on capturing nonlinear connections between the features.

B. Ensemble method implementation

The ensemble method is mainly based on the DV algorithm described in [10], where multiple regressors are combined together using local performance estimates to generate the final prediction. In particular, the algorithm is divided into two different phases, described below.

Learning phase. First, the hyperparameters of LASSO and RF were optimized in a cross-validation setup by performing an exhaustive search over a fixed grid of parameter values. Then, we computed the estimation error for each data record (EE_R) and each base learner in the training set, in another cross-validation setup, as:

$$EE_R = | \widehat{IB}_{lr} - IB_{OPTr} | \quad (2)$$

where $\widehat{IB}_{lr}$ is the output obtained from the l-th learner (fitted on the k-1 training folds of the cross-validation setup) applied to the r-th record of the k-th validation fold, and IB_{OPTr} is the IB_{OPT} related to the r-th record. Finally, the two base learners, having the optimal hyperparameters previously determined, were trained using the whole training set.

Application phase. The model trained in the previous phase is applied as follows. For each input record of the test set, similar data through the training records were searched using the k-nearest neighbors (k-NN) method, to define its local region. The similarity measure is based on the weighted Euclidean distance, in which weights are derived from the application of RReliefF (RRF) algorithm to the dataset [13] , [14]. This procedure was performed to differently weigh the variables, being some features more relevant than others for the definition of the local subspace, i.e., more related to the target variable. Hence, after having defined the local subarea, for each base learner we predicted the $\widehat{EE}_R$ of the test record by averaging the EE_R of the nearest neighbors. Once the $\widehat{EE}_R$ of both LASSO and RF was derived, the final output was computed by weighting more the prediction of the best performing model than the other, according to $\widehat{EE}_R$. In particular, the prediction related to the model with the best outcome within the local region was weighted three times more than the one having the lowest performance.

IV. Model Evaluation

A. Quantification of model performance

We evaluated the efficacy of the ensemble on the test set, both in terms of regression metrics, i.e., root mean squared error (RMSE) and mean absolute error (MAE), and within the simulated environment, where the estimated IB was applied as mealtime insulin amount. Then, we computed popular metrics related to glycemic control over the 6 hours postprandial window [15]: the BGRI, the percentage of time spent within the target BG range ($70 \leq BG \leq 180$ mg/dL), the time above this range and below this range, hereafter named as %TIR, %TAR and %TBR respectively. In addition, the assessment was performed for the single base learners (LASSO and RF), to verify the benefit resulting from the usage of the ensemble compared to the single models. Finally, as a reference, we included the SF and OPT within the in silico evaluation.

B. Model interpretation

Interpretability of machine learning models is a relevant feature, especially when dealing with a clinical application. However, interpretation of complex nonlinear models is far from trivial, and the use of ad-hoc interpretability methods is needed. Therefore, in this work, we applied the SHAP tool [16], i.e., a novel game theoretic approach to explain how much a given feature impacts on model prediction in comparison to the prediction obtained if that feature took some baseline value.

TABLE I. COMPARISON OF THE METRICS OBTAINED FROM THE IN SILICO EVALUATION FOR SF, LASSO, RF, DV AND OPT. MEDIAN AND INTERQUARTILE RANGES ARE REPORTED FOR TAR, TIR AND TBR [%] AND BGRI.

	SF	LASSO	RF	DV	OPT
TAR [%]	29.09 [13.30-37.95]	29.92 [11.91-39.06]	30.19 *[12.19-40.17]*	29.92 [11.91-39.33]	29.92 [12.18-39.33]
TIR [%]	60.11 [38.23-78.12]	64.54 [48.20-79.22]	65.65 [52.35-81.16]	65.65 [51.52-80.89]	68.14 [59.00-82.27]
TBR [%]	0 [0-28.53]	0 [0-13.57]	0 [0-3.04]	0 [0-6.09]	0 [0-0]
BGRI	9.93 [4.85-17.46]	9.09 [4.68-15.44]	8.81 [4.34-14.96]	8.79 [4.40-14.93]	8.23 [3.98-14.03]

V. RESULTS AND DISCUSSION

A. Goodness of fit

The ensemble method led to an improvement compared to the single models in terms of regression metrics. In particular, the RMSE and MAE resulted 0.87 U, 0.90 U and 0.62 U, 0.60 U for the base regressors LASSO and RF, respectively, while the DV method produced a RMSE of 0.84 U and a MAE of 0.56 U. Moreover, we found out that the most important subdomain that allows differentiating the performance of the base learners is given by SF. Indeed, within the local subdomain defined from medium SF values (from 8 to 15 U), the LASSO model produced better results compared to RF, while the latter model outperformed LASSO at the extremities of SF domain. The DV approach allowed the combination of the positive contributions of LASSO and RF, improving the outcome.

B. Assessment of the methodology in terms of glucose control

A representative simulated scenario related to the use of SF, DV, and OPT is shown in Fig. 1, in which the hyperglycemic event caused by SF is considerably reduced with the application of DV, approaching the optimal control. Table I reports the median and interquartile ranges of the distributions of TIR%, TAR% and TBR% for each considered method. Compared to SF, LASSO, RF, and DV improved glucose control. In particular, resulting %TIR is 60.1, 64.5, 65.6, and 65.6 respectively.

Furthermore, comparing LASSO with RF, the former resulted in a lower median %TAR, while, conversely, the latter outperformed LASSO in terms of %TBR, reaching a 75th of about 3% against 13.6%. Focusing on the results obtained using DV, it allowed to maintain a high postprandial %TIR, while reaching a trade-off in terms of %TAR and %TBR. This result was achieved thanks to the ability of the ensemble approach to select as major contributor for the final outcome the base learner having better performance within specific local domains.

C. Model interpretation

The plot in Fig. 2 reports the application of SHAP to the DV model. In particular, information on both feature importance and feature effect on the outcome are shown. Each dot of the plot represents the impact on model outcome for a given feature and record of the dataset, and has three characteristics: the color, the vertical and the horizontal location. The different colors represent the magnitude of the feature value (from low, in blue, to high, in red). The vertical location shows the feature importance, indeed the features are ranked in descending order based on their effect on the model prediction, while the horizontal location reports whether the impact of that value caused a higher or lower prediction. The most important feature resulted SF, in accordance with the results obtained in subsection V.A, followed by $\dot{BG}$, and CHO, which provided a positive contribution to the model outcome. Instead, lower values of CR, IOB and CF led to higher IB amounts, in accordance with the physiological meaning of these variables, unlike BW and Gt, that should have an opposite impact on the model output with respect to the one obtained in Fig. 2. The interpretability issue with BW and Gt variables was already encountered in our previous work [8], where the BW and Gt coefficients of the linear model were not in line with the physiological interpretation of these features. As explained in [8], this result may be due to the non-zero correlation among some variables of the dataset.

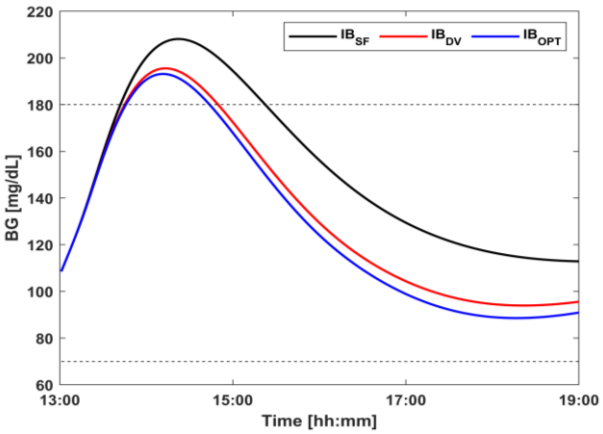

Fig. 1: Representative BG curves during the postprandial time window for different methods of insulin bolus computation (SF in black, DV in red and OPT in blue). Dashed lines indicated the target glycemic range.

VI. CONCLUSION

In this work, we developed an ensemble learning method based on dynamic voting with the aim of optimizing the mealtime IB in T1D therapy, thus preventing possible adverse glycemic events, such as hypo/hyperglycemia, due to an over/underestimation of IB amount. The proposed method was tested in a simulated environment, producing positive results in terms of trade-off between %TAR and %TBR,

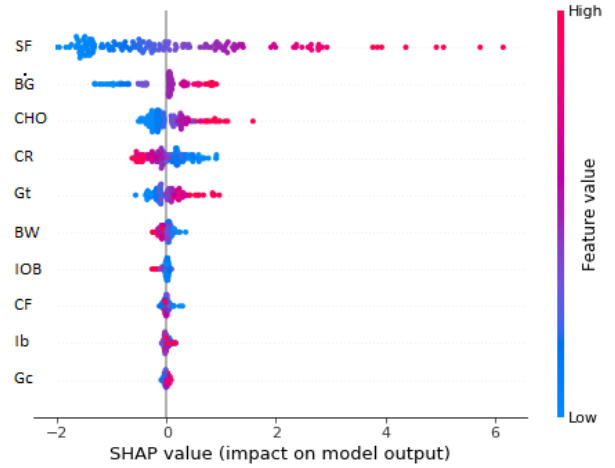

Fig. 2: Results of the interpretation of the DV model. Each point on the plot represents the impact on model output for a feature and a record of the dataset. The features are reported in descending order based on their importance within the model. Different colors of the points represent high/low feature values.

while keeping a high median %TIR.

Preliminary results showed the ability of the proposed model to combine the linear and nonlinear model of [8], [9] based on each individual regressor's local performance within a specific subdomain. Moreover, the model interpretation using SHAP confirmed the importance of information related to the glucose dynamics ($\dot{BG}$) within the model.

Limitations of the study are represented by the use of a synthetic dataset, which did not take into account confounding factors, such as error sources.

Further development of this work will include the assessment of the methodology in a multi-meal scenario [17], with the addition of variability and multiple sources of error, such as CGM measurement error [18], [19] and CHO counting error [20]. Moreover, future works will include the evaluation of the method using real data by leveraging a recently proposed in-silico framework to retrospectively assess new therapy guidelines for T1D management [21].

REFERENCES

[1] F. A. Karlsson *et al.*, "Beta-cell activity and destruction in type 1 diabetes.," *Ups. J. Med. Sci.*, vol. 105, no. 2, pp. 85–95, 2000.

[2] P. E. Cryer, S. N. Davis, and H. Shamoon, "Hypoglycemia in Diabetes," *Diabetes Care*, vol. 26, no. 6, pp. 1902–1912, 2003.

[3] S. Schmidt and K. Nørgaard, "Bolus Calculators," *J. Diabetes Sci. Technol.*, vol. 8, no. 5, pp. 1035–1041, 2014.

[4] P. Davidson, H. Hebblewhite, R. Steed, and B. Bode, "Analysis of Guidelines for Basal-Bolus Insulin Dosing: Basal Insulin, Correction Factor, and Carbohydrate-to-Insulin Ratio," *Endocr. Pr.*, vol. 14, no. 9, pp. 1095–1101, 2008.

[5] T. M. Gross, D. Kayne, A. King, C. Rother, and S. Juth, "A Bolus Calculator Is an Effective Means of Controlling Postprandial Glycemia in Patients on Insulin Pump Therapy," *Diabetes Technol. Ther.*, vol. 5, no. 3, pp. 365–369, 2003.

[6] D. De Salvo and B. Buckingham, "Continuous glucose monitoring: current use and future directions.," *Curr. Diab. Rep.*, vol. 13, no. 5, pp. 657–662, Oct. 2013.

[7] M. Vettoretti, G. Cappon, G. Acciaroli, A. Facchinetti, and G. Sparacino, "Continuous Glucose Monitoring: Current Use in Diabetes Management and Possible Future Applications," *J. Diabetes Sci. Technol.*, vol. 12, no. 5, pp. 1064–1071, 2018.

[8] G. Noaro, G. Cappon, M. Vettoretti, G. Sparacino, S. D. Favero, and A. Facchinetti, "Machine-Learning Based Model to Improve Insulin Bolus Calculation in Type 1 Diabetes Therapy," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 1, pp. 247–255, Jan. 2021.

[9] G. Noaro, G. Cappon, G. Sparacino, S. Del Favero, and A. Facchinetti, "Nonlinear Machine Learning Models for Insulin Bolus Estimation in Type 1 Diabetes Therapy," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, Jul. 2020, pp. 5502–5505..

[10] S. Puuronen, V. Terziyan, and A. Tsymbal, "A dynamic integration algorithm for an ensemble of classifiers," in *Foundations of Intelligent Systems*, 1999, pp. 592–600.

[11] C. Dalla Man, F. Micheletto, D. Lv, M. Breton, B. Kovatchev, and C. Cobelli, "The UVA/Padova Type 1 Diabetes Simulator: New Features," *J. Diabetes Sci. Technol.*, vol. 8, no. 1, pp. 26–34, 2014.

[12] W. Clarke and B. Kovatchev, "Statistical Tools to Analyze Continuous Glucose Monitor Data," *Diabetes Technol. Ther.*, vol. 11 Suppl 1, pp. S45-54, Jun. 2009.

[13] M. Robnik-Šikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, no. 1, pp. 23–69, 2003..

[14] J. Mendes-Moreira, A. M. Jorge, C. Soares, and J. F. de Sousa, "Ensemble Learning: A Study on Different Variants of the Dynamic Selection Approach," in *Machine Learning and Data Mining in Pattern Recognition*, 2009, pp. 191–205.

[15] T. Battelino *et al.*, "Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range.," *Diabetes Care*, vol. 42, no. 8, pp. 1593–1603, Aug. 2019.

[16] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, 2017, vol. 30, pp. 4765–4774.

[17] R. Visentin *et al.*, "The UVA/Padova Type 1 Diabetes Simulator Goes From Single Meal to Single Day," *J. Diabetes Sci. Technol.*, vol. 12, no. 2, pp. 273–281, 2018.

[18] A. Facchinetti, S. Del Favero, G. Sparacino, and C. Cobelli, "Model of glucose sensor error components: identification and assessment for new Dexcom G4 generation devices," *Med. Biol. Eng. Comput.*, vol. 53, pp. 1259–1269, Nov. 2014.

[19] M. Vettoretti, C. Battocchio, G. Sparacino, and A. Facchinetti, "Development of an Error Model for a Factory-Calibrated Continuous Glucose Monitoring Sensor with 10-Day Lifetime," *Sensors*, vol. 19, p. 5320, Dec. 2019.

[20] C. Roversi, M. Vettoretti, S. Del Favero, A. Facchinetti, and G. Sparacino, "Modeling Carbohydrate Counting Error in Type 1 Diabetes Management.," *Diabetes Technol. Ther.*, vol. 22, no. 10, pp. 749–759, Oct. 2020.

[21] G. Cappon et al. "ReplayBG: A Novel In-Silico Framework to Retrospectively Assess New Therapy Guidelines for Type 1 Diabetes Management", *Diabetes Technology Meeting*, Bethesda, MD, USA, 2020