Personalized Machine Learning Algorithm based on Shallow Network and Error Imputation Module for an Improved Blood Glucose Prediction

Jacopo Pavan, Francesco Prendin, Lorenzo Meneghetti, Giacomo Cappon, Giovanni Sparacino, Andrea Facchinetti, Simone Del Favero [1]

Abstract. Real-time forecasting of blood glucose (BG) levels has the potential to drastically improve management of Type 1 Diabetes, a widespread chronic disease affecting the metabolic system. Most notably, if hypo or hyperglycemia episodes (i.e. glycemic excursion below or above a safe range) could be accurately predicted, then the patient could be timely warned, thus enabling proactive countermeasures to avoid these dangerous conditions. In this work, a novel personalized algorithm for the real-time forecasting of BG is developed by combining the output of a shallow feed forward neural network with an error imputation module composed by an ensemble of trees. Past glucose readings as well as insulin, meals and work/sleep time information are carefully handled to train and boost the prediction performance of the algorithm. The root mean square error over the 6 subjects achieves a mean value of 18.69 mg/dL and 32.43 mg/dL for 30- and 60-minute prediction horizon respectively.

1 INTRODUCTION

Type 1 diabetes (T1D) is a metabolic disease characterized by the destruction of the pancreatic cells responsible for insulin production and thus resulting in an impaired Blood Glucose (BG) homeostasis. Diabetes treatment relies on external insulin injection aimed at maintaining BG levels within a physiological range [1], since poor BG regulation is responsible of comorbidities and reduced life expectancy [2]. In particular, concentrations below or above the normal range (called hypo- and hyper-glycemia respectively) can either represent an immediate threat to patient safety or could cause severe long-term complications.

Measures to mitigate/avoid these conditions include the administration of exogenous insulin to reduce hyperglycemic excursion or the consumption of fast-acting carbohydrates for hypoglycemic events. Unfortunately, both insulin injection and carbohydrates consumption affects BG only after a considerable amount of time: 45 minutes for insulin and at least 15 minutes for carbohydrates.

Therefore, accurate BG prediction would be of paramount importance to enable timely corrective actions and thus ensure successful BG control. This holds true both for standard therapy, where corrective actions are manually performed by patients, and in automated and semi-automated systems such as an artificial pancreas [1].

Nevertheless, BG prediction is by no mean a trivial task. BG concentration is the result of complex non-linear dynamical interaction of multiple physiologic subsystems and is influenced by several factors, often hard to measure. They include timing and magnitude of carbohydrates consumption, protein and fat content of the meal, length, intensity and even type (aerobic vs anaerobic) of physical exercise, stress, illness or menstrual cycle. Furthermore, the modeling of BG dynamics is hindered by the large variability in the physiological metabolic response of different individuals. For these reasons, non-linear and personalized models can represent one of the best options to address the task. The introduction of continuous glucose monitoring (CGM) sensors in T1D care has encouraged the use of data-driven models based on past BG readings, whereas the further availability of data offered by infusion pumps and fitness bands opened the possibility of using exogenous inputs as additional features for describing the model [15, 3].

Over the last two decades, several non-linear algorithms have been tested in this framework, including support vector machine [4], gaussian process regression [14], random forest (RF) [5] and several kinds of neural networks (NN) [17, 10, 16, 9], up to deep-learning approaches like long short-term memory networks [13] and convolutional NN [6]. While some works assessed that only past CGM information is actually useful to describe an accurate model [6, 13, 8], many others claims that exogenous inputs play an important role in describing BG dynamics. These extra features include time of the day, insulin administration, food intake, energy expenditure, lifestyle and emotions [5, 4, 6, 10, 18].

Up to the present, none of the these models has stand out from the others in terms of prediction accuracy. This consideration lead our group to focus more on feature manipulation, selection and hyperparameters optimization and to explore the possibility of combining different kinds of non-linear learners. In conclusion, the aim of our work is to synthesize an accurate model of BG dynamics by starting from a simple, feedforward NN and to investigate:

- the impact of hyperparameters optimization and feature selection;
- the improvement achievable by combining the NN with an error imputation module (EIM) based on a regression trees ensemble.

2 DATASET

Our study is based on the real patient data provided by the OhioT1DM Dataset, described in detail in [7]. In its 2020 update, six new patients are introduced in the dataset, with roughly 8 weeks of data each (6 weeks of training set, 10 days of testing set). Each of them wore a CGM device (Medtronic Enlite CGM sensor), and an insulin pump (Medtronic 530G or 630G). Daily life-events (e.g.

[1] Department of Information Engineering, University of Padova, Italy, email: {pavanjac, prendinf, meneghet, cappongi, gianni, facchine, sdelfave}@dei.unipd.it

work/sleep, exercise) are reported via smartphone app while other physiological data (e.g. skin temperature) are provided by using Empatica Embrace fitness wristbands. We will work exclusively on these six new subjects.

Some of these signals are quasi-continuously measured, like CGM, acceleration or basal insulin. Some others are impulse-like, e.g. self-monitoring BG or meal consumption, which are provided only a few number of times along the day. Some others, like work intensity or sleep quality, are instead defined for time windows of the order of some hours.

Some of the impulsive-like features, e.g. insulin boli, have an impact on glucose dynamics that could last up for several hours. In order to include this information in the framework of feedforward NNs, which have no memory about the dynamics of inputs, we had to rely on new features. We described the insulin-on-board (IOB) by convolution of insulin boli and basal with a 6 hours activity curve, whereas the convolution of consumed carbohydrates with an absorption curve (different for slow- and fast-acting carbohydrates) returned the carbohydrates-on-board (COB). These two variables carry information about the dynamics of slow insulin absorption and carbohydrates slow impact on BG, hence they are suitable for being used with feedforward NNs. In a similar way, we described the physical-exercise-on-board by low-pass filtering with second order transfer function the physical activity intensity.

We also introduced the slope of CGM, computed by using the last 2 hours of readings, since the trend of the glycemic profile resulted to be significantly correlated to future CGM readings in the training set. Other tested features include daytime, time and amount of last carbohydrates intake and several filtered version of the original signals.

3 METHODS

As shown in Figure 1, the proposed model is composed by two parts. The first one is a shallow NN, which is the main predictor. It is trained to predict future BG values with a certain prediction horizon (PH). The second predictor is based on an ensemble of trees. It is called error imputation module, since it is trained to predict the error that is committed by the shallow NN. Finally, as shown in Figure 1, the prediction of the proposed algorithm is obtained by combining the output of the shallow neural network $CGM_s(t+PH)$ with the output of EIM, $\hat{e}(t+PH)$, to have an accurate value for the expected glucose concentration.

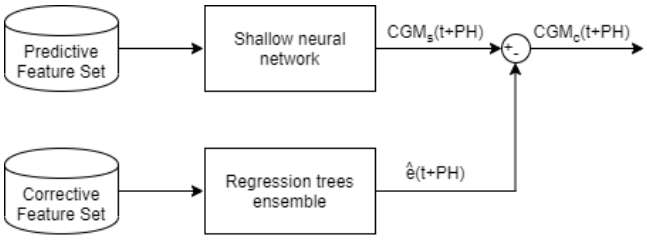


Figure 1: Complete architecture of the predictive model. Variable $CGM_s(t+PH)$ is the original prediction, $CGM_c(t+PH)$ is the corrected prediction and $\hat{e}(t+PH)$ is the predicted error.

3.1 SHALLOW NEURAL NETWORK

The feature set for the shallow NN is manually determined on a population level, i.e. by looking for a unique set of feature which can be used by all subjects. The first criterion for feature selection consists in excluding all those features that presents too many missing values in the training set and hence cannot be considered reliable. For instance, COB was discarded because, in some subjects, the information about meal is missing for a large part of the training set (e.g. subject 567 reported only 31 of the expected 141 meals). A second criterion consists in excluding the features that are expected to have a negligible impact on prediction accuracy, e.g. the state of illness or work. This selection was performed with domain experts and by means of some preliminary evaluation on the training set. The selected features still presented many missing values that could hinder the training procedure. Thus we performed, exclusively on the training set, a first order interpolation on any gap of samples shorter than 30 minutes. Finally, we investigated how many past CGM readings should be used as inputs in order to improve model accuracy. To this purpose we performed a bayesian optimization [12], using 70% of the data in the original training set for training and the remaining 30% for validation. The result was that using the last 16 CGM instants (corresponding to the last 1 hour and 20 minutes of readings) leads to the optimal prediction performances in the six subjects. In conclusion, the resulting feature pool we adopted included present and past CGM readings, CGM slope and IOB.

Similarly, we determined the number of hidden layers and number of neurons in each layer with an exploratory optimization, which was also performed via bayesian optimization and using the same training/validation split we employed for feature selection. The best performances on the validation set were found when using a single hidden layer with a reduced number of inner nodes (only 5).

While the architecture of the net was optimized on a population level and it will be the same for every subject, the weights of the net were trained individually for every subject, meaning that the resulting model is tailored on each of them. The shallow net is trained on the whole original training set, with its target being either the 6- or 12-steps ahead prediction of CGM, i.e. prediction horizons of 30 and 60 minutes, respectively. Inputs and targets of the net are normalized by the mean and standard deviation computed on the training set.

3.2 ERROR IMPUTATION MODULE

We noticed that the prediction of the main NN is affected by a large error when abrupt changes occur in its inputs. For instance, consecutive CGM readings showing a large difference in values are often associated with poor BG predictions. However, other explanatory factors for the prediction error could be found in those features which were not used by the shallow NN. The key idea behind the EIM is to provide an estimate $\hat{e}(t+PH)$ of the true error, $e(t+PH)$, affecting the prediction of the shallow net. The first step to build this module is to create a new pool of feature, named Corrective Feature Set, containing the first order differences, at several time lags, of: CGM, IOB, COB, sleep/work period, skin temperature and acceleration data. Then, to take into account the large inter-individual variability, a feature selection step (based on ReliefF [11]) is applied to the Corrective Feature Set. Then, only the features with highest ranks are used to train the model. A further level of individualization is achieved by optimizing several hyperparameters. For each subject, a bayesian optimization procedure returns the best method to train the ensemble of trees (i.e., Bagging or Boosting), the best number of trees (searched among the range {10,500}), the best number of leaves for each tree (searched among the log-scaled range {1,max(2,number of training sample/2)}), the best tree-depth (searched among the log-scaled range {1,max(2,number of train-

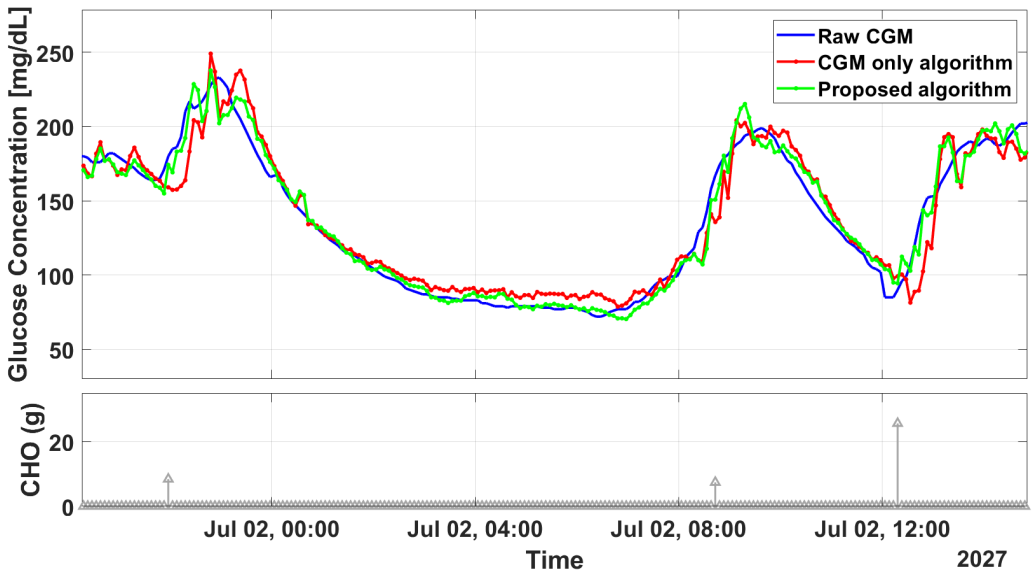


Figure 2: 30-minute prediction for subject 544. The three curves represent real CGM data (blue line), the prediction with CGM-NN (red line), and the prediction with NN-EIM (green line).

ing sample-1)}), the best learning rate (among the log-scaled range $\{0.001,1\}$ if boosting method is chosen). Both the feature selection and the hyperparameters optimization exploits the training set only.

3.3 BENCHMARK NEURAL NETWORK

The effectiveness of the proposed approach is assessed by comparing the predicted profiles with the ones obtained by a benchmark predictor: a shallow neural network based on CGM data only (CGM-NN). This network resorts the structure of the main predictor (i.e. a single hidden layer with 5 inner nodes) but it exploits past CGM data only (16 samples) as input features. This model will also be personalized for each patient.

4 METRICS

The accuracy of the predicted profiles is evaluated by using four metrics. The Mean Absolute Error (MAE) is defined as:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}(y(t) - \hat{y}(t|t-PH))$$

where PH is the prediction horizon, $y(t)$ is the current CGM reading, N is the length of the whole signal y and $\hat{y}(t|t-PH)$ is the the PH-steps ahead prediction using the information available up to instant t. Similarly we can define the Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y(t) - \hat{y}(t|t-PH))^2}$$

and Coefficient of Determination (COD):

$$COD = 100 \cdot (1 - \frac{||(y(t) - \hat{y}(t|t-PH))||_2^2}{||(y(t) - \bar{y}(t))||_2^2})$$

where $\bar{y}(t)$ is the average value of y. COD counts for the variance explained by the predictive model with respect to the total variance of the signal. Its maximum value is 100%.

The delay existing between the target signal and the predicted one can be computed as the temporal shift that minimizes the square of the mean quadratic error between these two signals:

$$delay = \underset{j\in[0,PH]}{\arg\min}\Big[\frac{1}{N}\sum_{t=1}^{N-PH}((\hat{y}(t|t-PH)+j) - y(t))^2\Big].$$

5 RESULTS

Since the shallow net is initialized with random weights, results will be reported in terms of mean and standard deviation of the various metrics on 10 different initialization of the algorithm.

We compared two models: one is the benchmark, shallow NN using exclusively past and present CGM readings (CGM-NN); the other is the shallow net employing the Predictive Feature Set and the error imputation module (NN-EIM). Table 1 and Table 2 reports the results obtained with these two models, on each subject, for 30 and 60 minutes prediction horizons respectively. The last row of the tables averages the mean values of the metrics on every subject.

As aforementioned in Section 3.1, no kind of operation was performed on the testing set in order to impute missing values. Table 3 reports the number of CGM samples available for each subject and the number of those predicted by our CGM-NN and NN-EIM.

NN-EIM achieves better results on each subject for all the evaluation metrics, both for PH = 30 and PH = 60 minutes. On average, the RMSE improves from 19.50 mg/dL with CGM-NN to 18.63 mg/dL with NN-EIM on the 30 minutes PH (p-value=0.031) and from 34.26 mg/dL with CGM-NN to 32.27 mg/dL with NN-EIM (p-value=0.031). The p-values are computed with a Wilcoxon signed rank test and show that the improvement, albeit small ($\sim 5\%$ in both cases), is statistically significant with $1-\alpha = 0.95$ confidence level.

Figure 3 show the boxplots and scatter plots of the average RMSE values for every subject, for a PH of 30 minutes (Figure 3a) and 60 minutes (Figure 3b). The color of the lines linking the scatter plots indicates the magnitude of the difference in RMSE between the two strategies: green-shaded lines mean an improved accuracy from CGM-NN to NN-EIM; viceversa, the lines are red if model accuracy worsen. Both for 30 and 60 minutes ahead predictions, the use

Table 1: Evaluation of RMSE, MAE, COD and delay (mean (± standard deviation)) with CGM-NN and NN-EIM on a 30 minutes PH.

Subj	RMSE		MAE		COD		delay	
	CGM-NN	NN-EIM	CGM-NN	NN-EIM	CGM-NN	NN-EIM	CGM-NN	NN-EIM
540	21.66 (±0.20)	20.42 (±0.27)	14.09 (±0.13)	12.81 (±0.14)	90.29 (±0.18)	91.58 (±0.22)	20 (±0.0)	15.0 (±0)
544	17.67 (±0.08)	16.50 (±0.22)	10.52 (±0.09)	9.57 (±0.13)	89.86 (±0.10)	91.33 (±0.24)	20.0 (±0.0)	15.0 (±0)
552	16.81 (±0.15)	16.51 (±0.12)	6.85 (±0.13)	6.51 (±0.07)	90.72 (±0.16)	91.38 (±0.12)	20.0 (±0.0)	17.5 (±2.6)
567	20.73 (±0.10)	20.18 (±0.14)	10.92 (±0.09)	10.19 (±0.08)	86.89 (±0.13)	88.14 (±0.16)	20.0 (±0.0)	17.5 (±2.6)
584	22.44 (±0.14)	21.29 (±0.33)	12.77 (±0.14)	11.34 (±0.18)	87.69 (±0.15)	89.71 (±0.32)	25.0 (±0.0)	25.0 (±0.0)
596	17.71 (±0.29)	16.89 (±0.24)	10.93 (±0.08)	10.09 (±0.15)	88.28 (±0.38)	89.53 (±0.30)	25.0 (±0.0)	20.0 (±0.0)
All	19.50 (±2.39)	18.63 (±2.22)	11.01 (±2.45)	10.08 (±2.10)	88.95 (±1.54)	90.28 (±1.37)	21.6 (±2.5)	18.3 (±3.7)

Table 2: Evaluation of RMSE, MAE, COD and delay (mean (± standard deviation)) with CGM-NN and EIM-NN on a 60 minutes PH.

Subj	RMSE		MAE		COD		delay	
	CGM-NN	NN-EIM	CGM-NN	NN-EIM	CGM-NN	NN-EIM	CGM-NN	NN-EIM
540	40.34 (±0.82)	37.69 (±0.60)	26.46 (±0.51)	23.83 (±0.36)	66.61 (±1.36)	71.53 (±0.91)	45.0 (±0.0)	40.0 (±0)
544	31.21 (±0.31)	28.74 (±0.32)	19.57 (±0.30)	16.61 (±0.21)	68.55 (±0.64)	73.90 (±0.60)	44.5 (±1.5)	27.0 (±2.58)
552	30.45 (±0.17)	29.56 (±0.23)	12.56 (±0.14)	11.89 (±0.18)	70.12 (±0.33)	72.93 (±0.43)	45.0 (±0.0)	40.0 (±0.0)
567	37.37 (±0.29)	36.28 (±0.40)	20.48 (±0.24)	18.65 (±0.12)	67.13 (±0.65)	62.70 (±0.83)	45.0 (±0.0)	41.0 (±2.1)
584	36.85 (±0.27)	33.84 (±0.26)	21.26 (±0.32)	18.37 (±0.18)	67.13 (±0.49)	74.52 (±0.40)	53.0 (±2.4)	46.0 (±2.1)
596	29.32 (±0.23)	27.51 (±0.57)	18.58 (±0.20)	16.74 (±0.25)	68.31 (±0.51)	72.53 (±1.15)	49.0 (±2.1)	42.5 (±2.6)
All	34.26 (±4.5)	32.27 (±4.25)	19.82 (±4.49)	17.69 (±3.87)	66.55 (±4.08)	71.35 (±4.36)	47.0 (±3.5)	39.4 (±6.4)

Table 3: Number of CGM samples available in the testing set of the OhioT1DM Dataset 2020 and number of CGM samples predicted by CGM-NN and NN-EIM.

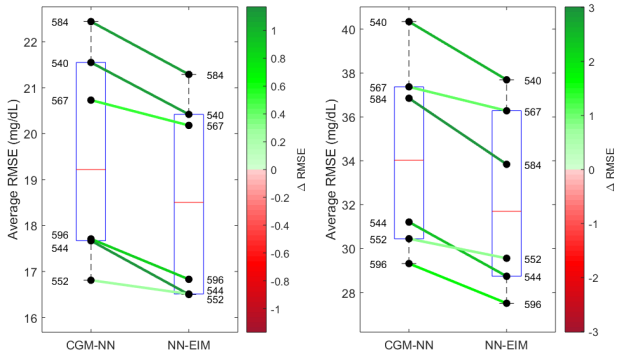
Subj	Available samples	Predicted samples	
		PH=30	PH=60
540	2884	2610	2592
544	2704	2532	2514
552	2352	2061	2020
567	2377	2050	1996
584	2653	2206	2165
596	2731	2551	2521

of exogenous inputs and EIM results in a systematic improvement of performances.

The RMSE of the NN-EIM ranges from 16.50 mg/dL up to 21.29 mg/dL with PH=30 for subject 544 and 584, respectively. This difference might be related to the occurrence of large oscillations in CGM data which cannot be clearly explained by any input variable in the dataset. Considering a PH = 30 min, the lowest improvement in terms of RMSE is for subject 552 (16,81 mg/dL vs 16.51 mg/dL, CGM-NN vs NN-EIM respectively). The largest one is for subject 540 (21.66 mg/dL vs 20.42 mg/dL, CGM-NN vs NN-EIM respectively). One of the main reasons linked to this marginal improvement is the large number of missing values in the testing set for some features (e.g. acceleration and skin temperature).

Nevertheless, whenever the feature set is reliable — as for subject 544 in Figure 2 — the proposed algorithm (green line) reduces the prediction delay as well as the error when CGM data show a positive increment related to any external input (i.e. CHO ingested, in this case). Furthermore, the use of lagged first order differences of past CGM data provides a prediction which is more adherent to the target CGM than the one obtained by CGM-NN. This is confirmed by an enhanced COD for subject 552 (90,72% vs 91,38%, CGM-NN vs NN-EIM and by the delay (20 min vs 17,5 min, CGM-NN vs NN-EIM).

Same conclusions can be found by considering a PH = 60 min. As before, the lowest improvement in terms of RMSE is for subject 552 (30,45 mg/dL vs 29,56 mg/dL, CGM-NN vs NN-EIM). The largest is for subject 584 (36,85 mg/dL vs 33,84 mg/dL, CGM-NN vs NN-EIM). The prediction capabilities of the proposed approach are also confirmed by COD and delay, even when a marginal improvement in terms of RMSE is found. In fact, for subject 552 we found the COD for CGM only NN vs proposed algorithm is 70,12% vs 72,93%. Delay is reduced at 40 minutes, by meaning that the prediction has a useful time anticipation of 20 minutes.



(a) PH=30 minutes (b) PH=60 minutes

Figure 3: Boxplot and scatter plots of average RMSE obtained with CGM-NN and EIM-NN on a 30 minutes (a) and 60 minutes (b) prediction horizons.

6 CONCLUSIONS

In this work we presented a new approach for a real-time forecasting of glucose levels based on a shallow neural network and an error imputation module (NN-EIM). Comparing the performance, at PH = 30 and PH = 60, of the novel algorithm vs CGM-NN, we demonstrated that accurate feature manipulation and selection steps can effectively improve the prediction accuracy and reduce the delay affecting the

predicted profile. However, the presence of many missing values in some variables reduce the improvement brought by the proposed approach. Finally, a further development of this work is the investigation of patterns within the CGM time series and the inclusion of such physiological priors into the proposed algorithm could results in improved performance.

ACKNOWLEDGMENTS

This work was partially supported by MIUR, (Italian Minister for Education, under the initiatives "Departments of Excellence" (Law 232/2016) and "SIR: Scientific Independence of young Researchers", project RBSI14JYM2 "Learn4AP: Patient-Specific Models for an Adaptive, Fault-Tolerant Artificial Pancreas".

CODE

A repository containing the code developed for this work is available at: `https://github.com/jp993/BGPC_2020`

REFERENCES

[1] C. Cobelli, C. Dalla Man, G. Sparacino, L. Magni, G. De Nicolao, and B. Kovatchev, 'Diabetes: Models, signals and control', *IEEE Rev Biomed*, **2**, 54–96, (2009).

[2] L.A DiMeglio, C. Evans-Molina, and R. Oram, 'Type 1 diabetes', *Lancet*, **391**, 2449–62, (2018).

[3] A. Gani, A.V. Gribok, S. Rajaraman, W.K. Ward, and J. Reifman, 'Predicting subcutaneous glucose concentration in humans: data-driven glucose modeling', *IEEE Trans Biomed Eng*, **56**, 246–54, (2009).

[4] E. Georga, V.C. Protopappas, D. Ardigo, M. Marina, I. Zavaroni, D. Polyzos, and D.I. Fotiadis, 'Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression', *IEEE J Biomed Health Inform*, **17**, 71–81, (2013).

[5] E. Georga, V.C. Protopappas, D. Polyzos, and D. Fotiadis, 'A predictive model of subcutaneous glucose concentration in type 1 diabetes based on random forest', *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2889–92, (2012).

[6] K. Li, J. Daniels, C. Liu, P. Herrero, and P. Georgiou, 'Convolutional recurrent neural networks for glucose prediction', *IEEE J Biomed Health Inform*, **24**, 603–613, (2019).

[7] C Marling and R. Burnescu, 'The ohiot1dm dataset for blood glucose level prediction: Update 2020', (2019).

[8] J. Martinsson, A. Schliep, B. Eliasson, C. Meijner, S. Persson, and O. Mogren, 'Automatic blood glucose prediction with confidence using recurrent neural networks', *Proceedings of the 3rd International Workshop on Knowledge Discovery in Healthcare Data*, 64–68, (2018).

[9] S.M. Pappada, B.D. Cameron, and P.M. Rosman, 'Development of a neural network for prediction of glucose concentration in type 1 diabetes patients', *J Diabetes Sci Technol*, **2**, 792–801, (2008).

[10] C. Perez-Gandia, A. Facchinetti, G. Sparacino, C. Cobelli, E.J. Gḿez, M. Rigla, A. de Leiva, and M.E. Hernando, 'Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring', *Diabetes Technol Ther*, **12**, 81–88, (2010).

[11] M. Robnik-Sikonha and I. Kononenko, 'Theoretical and empirical analysis of relieff and rrelieff', *Machine Learning*, **53**, 23–69, (2003).

[12] J. Snoek, H. Larochelle, and R.P. Adams, 'Practical bayesian optimization of machine learning algorithms', *Advances in Neural Information Processing Systems*, **4**, 2951–2959, (2012).

[13] Q. Sun, M.V. Jankovic, L. Bally, and S. G. Mougiakakou, 'Predicting blood glucose with an lstm and Bi-LSTM based deep neural network', *2018 14th Symposium on Neural Networks and Applications (NEUREL)*, 1–5, (2018).

[14] J. Valletta, A. Chipperfield, and C. Byrne, 'Gaussian process modelling of blood glucose response to free-living physical activity data in people with type 1 diabetes', *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4913–6., (2009).

[15] A.Z. Woldaregay, E. Å rsand, W. Stå le, A. David, L. Mamykina, T. Botsis, and G. Hartvigsen, 'Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes', *Artificial Intelligence in Medicine*, **98**, 109–134, (2019).

[16] Z. Zainuddin, O. Pauline, and C. Ardil, 'A neural network approach in predicting the blood glucose level for diabetic patients', *Int J Comput Intell*, **5**, 72–79, (2009).

[17] C. Zecchin, A. Facchinetti, G. Sparacino, and C. Cobelli, 'Jump neural network for online short-time prediction of blood glucose from continuous monitoring sensors and meal information', *Comput Methods Programs Biomed*, **113**, 144–52, (2013).

[18] C. Zecchin, A. Facchinetti, G. Sparacino, and C. Cobelli, 'How much is Short-Term glucose prediction in type 1 diabetes improved by adding insulin delivery and meal content information to cgm data? a Proof-of-Concept study', *J Diabetes Sci Technol*, **10**, 1149–60, (2016).