

Predicting Quality of Overnight Glycaemic Control in Type 1 Diabetes using Binary Classifiers

Amparo Güemes, Giacomo Cappon, Bernard Hernandez, Monika Reddy, Nick Oliver, Pantelis Georgiou, *Senior Member, IEEE*, Pau Herrero

Abstract—In type 1 diabetes management, maintaining nocturnal blood glucose within target range can be challenging. Although semi-automatic systems to modulate insulin pump delivery, such as low-glucose insulin suspension and the artificial pancreas, are starting to become a reality, their elevated cost and performance below user expectations is hindering their adoption. Hence, a decision support system that helps people with type 1 diabetes, on multiple daily injections or insulin pump therapy, to avoid undesirable overnight blood glucose fluctuations (hyper- or hypoglycaemic) is an attractive alternative. In this paper, we introduce a novel data-driven approach to predict the quality of overnight glycaemic control in people with type 1 diabetes by analyzing commonly gathered data during the day-time period (continuous glucose monitoring data, meal intake and insulin boluses). The proposed approach is able to predict whether overnight blood glucose concentrations are going to remain within or outside the target range, and therefore allows the user to take the appropriate preventive action (snack or change in basal insulin). For this purpose, a number of popular established machine learning algorithms for binary classification were evaluated and compared on a publicly available clinical dataset (i.e. OhioT1DM). Although there is no clearly superior classification algorithm, this study indicates that, by using commonly gathered data in type 1 diabetes management, it is possible to predict the quality of overnight glycaemic control with reasonable accuracy (AUC-ROC=0.7).

Index Terms—Decision support systems, machine learning, glycaemic control, night quality, type 1 diabetes

I. INTRODUCTION

TYPE 1 diabetes (T1D) is a long-term condition characterized by loss of insulin secretion by the pancreatic β cells [1]. Standard insulin therapy is very demanding for people living with T1D. In particular, people with T1D measure capillary blood glucose several times a day and administer exogenous insulin via multiple daily injections (MDI) or

continuous subcutaneous insulin infusion (CSII). Continuous glucose monitoring (CGM) technology [2] has enabled more advanced technologies to support self-management such as sensor-augmented insulin pumps with low-glucose insulin suspend [3], the artificial pancreas [4], and decision support systems (DSS) for insulin dosing [5]. Although the artificial pancreas might be, a priori, the holy grail of the technological solutions for glucose management, slow market adoption, elevated cost, and some unmet customer expectations [6], may limit uptake and emphasize the importance of research focused on DSS [7].

DSSs are software tools designed to help people with diabetes to improve blood glucose (BG) control in their daily routine. For instance, typical a DSS consist of alerts notifying the user of potential future adverse events, such as hypoglycaemia and hyperglycaemia [8]. They might also suggest the administration of meal insulin boluses or corrective insulin boluses to mitigate hyperglycaemia [9,10,11], recommend the intake of carbohydrates (CHO) to tackle hypoglycaemia (rescue CHO) [12, 13, 14], or provide suggestions to prevent exercise-induced hypoglycaemia [15].

Another less studied application where DSS for T1D management can be very useful is the provision of recommendations to improve overnight glycaemic control [16]. There is significant clinical evidence that overnight glycaemic control is affected by the behaviour of the person with diabetes during the day [17]. Therefore, a DSS that recommends measures, such as adjusting overnight basal insulin dosing, or whether rescue CHO should be taken before going to bed, could be a very useful tool.

A significant amount of research has been done for forecasting blood glucose levels within a short to mid-term horizon (15min-2hours) [18], and some of this work has been translated to commercial products, such as the predictive low-glucose insulin suspension systems (Medtronic 640G and Tandem Basal-IQ). However, such short prediction horizons might not always be sufficient to prevent overnight

A. Güemes, P. Georgiou, P. Herrero, B. Hernandez are with Centre for Bio-inspired Technology, Imperial College London, London, United Kingdom. (e-mail: {amparo.guemes15, pantelis, p.herrero-vinias, b.hernandez-perez}@imperial.ac.uk).

G. Cappon is with the Department of Information Engineering (DEI), University of Padova, Padova, Italy (e-mail: cappongi@dei.unipd.it).

M. Reddy, N. Oliver are with the Department of Medicine, Imperial College London, London, United Kingdom. (e-mail: {m.reddy, nick.oliver}@imperial.ac.uk).

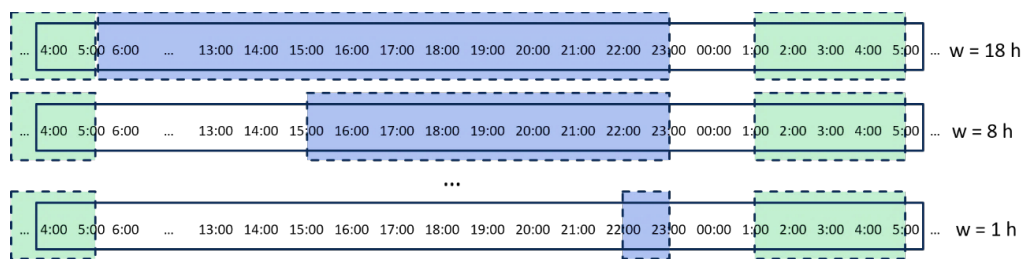


Fig. 1. Segmentation of daily profiles into day and night periods. The blue boxes represent the temporal windows used to extract the day-time features (e.g. 1h, 8h or 18h windows). The green boxes represent the temporal windows used to extract the night-time features. Note that night-time features are used to compute the category label for three independent binary classification problems: night-time within glucose range, with hypoglycaemia and with hyperglycaemia)

hypoglycaemia and do not prevent the user from waking up during the night in the case of a prediction alert being triggered.

Machine learning techniques are becoming increasingly popular to solve many T1D management problems, such as glucose forecasting [19], optimal insulin dosing [9,10], patient risk stratification [20], and CGM fault detection [21], as a result of their ability to represent complex non-linear input-output relationships, such as glucose-insulin dynamics. They can also be used to classify the quality of glycaemic control. A preliminary work on this subject has been proposed by Bertachi et al. looking at predicting nocturnal hypoglycaemia using artificial neural networks [22]. However, this work just focuses on hypoglycaemia prediction and evaluates only one classification algorithm.

In this work, we hypothesise that it is possible to predict the quality of overnight glycaemic control of a person with T1D by analysing the features extracted from CGM and commonly user-reported data (CHO and insulin) from a predefined temporal window within the preceding day-time period. In particular, three independent binary classification problems are defined to determine, at bedtime, the probabilities of three metrics associated with the quality of overnight glycaemic control: 1) the percentage time spent in target range, 2) the presence of nocturnal hypoglycaemia, and 3) the presence of nocturnal hyperglycaemia. To do so, different commonly used machine learning models (binary classifiers) are evaluated on a clinical dataset and their performance is assessed in terms of their classification accuracy. Finally, the best strategies, and their possible integration in a DSS, are discussed.

II. METHODS

A. Data

The publicly-available OhioT1DM dataset was used in this study to evaluate the proposed approach [23]. This dataset contains 8 weeks of data corresponding to six participants with T1D between 40 and 60 years old, all of them on insulin pump therapy and CGM. During the 8 weeks of study, participants were using the Medtronic 530G insulin pumps and the Medtronic Enlite CGM sensors (Medtronic Diabetes, Northridge, CA, US). Participants also reported life-event data, such as mealtime and CHO intake, using a smartphone app. Finally, physiological data was collected from a Basis Peak fitness band (Intel, Santa Clara, CA, US). Among all the data available in this dataset, the following physiological data were used in this study: CGM blood glucose level sampled every five minutes; insulin doses, both bolus and basal; user self-reported

meal times with carbohydrate estimates; time and amount of rescue carbs, and time of self-reported hypoglycaemic events.

B. Pre-processing of Raw Data

Despite robust data collection methodology, the OhioT1DM dataset still contains some data corruption, mainly due to erroneous manual inputs and missing data. Therefore, the following steps were followed to clean the raw data:

- Missing data from the CGM, which might happen due to sensor failure or data transmission problems, was addressed by discarding days with two consecutive hours of missing data. Otherwise, missing data was interpolated using the spline method. Moreover, the first and last day for each patient were also discarded as they were frequently incomplete.
- Outliers in the user self-reported CHO content estimation were identified and corrected (e.g. three times beyond the standard deviation). These outliers were replaced by the mean of the CHO intake over the 8 weeks of data collection.

In addition, to reduce the errors arising from using the discrete time series of CHO intake and insulin boluses, the following data transformations based on physiological models were performed to convert them into continuous time series. CHO intake was converted to the glucose rate of appearance into the systemic circulation using the gastrointestinal absorption model developed by Hovorka et al. [24]. Moreover, plasma insulin concentration due to subcutaneous insulin infusion was estimated using a validated model of subcutaneous insulin absorption [24]. Finally, data was segmented as follows. The 8-weeks continuous profiles were divided into individual days, each of them starting at 5am. This time was selected after visual inspection of the daily profiles, concluding that the majority of participants woke up between 5am to 6am. Then, these daily-profiles were further divided into day-time and night-time. Night-time was defined from 1am to 5am in order to exclude the impact of the postprandial excursion after dinner and the breakfast time. The period defined as day-time was chosen by using windows of different lengths (1h, 3h, 5h, 8h, 12h and 18h) before 11pm (see Fig. 1). The period from 11pm to 1am was not considered for feature extraction as it coincides with the postprandial excursion after dinner.

TABLE I
CALCULATED FEATURES FROM THE CGM, MEAL INTAKE, AND INSULIN DOSAGE DATA

Category	Feature	Abbreviation
Temporal domain	Mean CGM	mean_cgm
	Standard deviation CGM	std_cgm
	Risk index	ri
	Low blood glucose index	lbgi
	High blood glucose index	hbgi
	Percentage of time in target (70 -180 mg/dl)	time_in
	Percentage of time below target	time_under
	Percentage of time above target	time_above
Frequency domain	Four largest Fourier coefficients	P_1, P_2, P_3, P_4
	Frequencies corresponding to largest coefficients	f_1, f_2, f_3, f_4
Insulin dosage	Mean plasma insulin concentration	ins
Meal intake	Mean CHO rate of rate of glucose appearance (Ra)	
Hypoglycaemic events	Total amount of rescue CHO	hypo_correc

C. Feature Extraction and Labeling

For each day-time and night-time periods, a set of 19 features were extracted (see Table I). This set of input features includes: i) 8 indices to describe glucose variability in the time domain; ii) 8 indices to describe glucose variability in the frequency domain (Fast Fourier Transform, FFT), which comprise the 4 highest amplitudes and their corresponding frequencies (discarding the DC component); and iii) features related to meal intake, insulin dosage and presence of self-reported hypoglycaemic events. Note that features extracted from the frequency domain of the CGM signal contain valuable information that contributes to acquiring a better characterization of the signal [25]. A sensitivity analysis was carried out in a preliminary study to assess the impact of the features extracted from the FFT on the overnight glycaemic control quality and the indices that demonstrated the greatest impact were incorporated into the features set. As a result, the complete set of input features are deemed to provide enough information to evaluate the risk of suffering severe glucose fluctuations, i.e. hypoglycaemia or hyperglycaemia, both during day and night times. Note that other features such as exercise, stress, and illness might also significantly affect glycaemic control. However, in this work, we hypothesise that

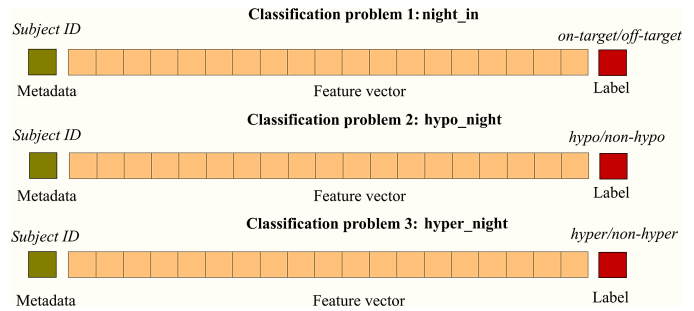


Fig. 2. Description of each observation: metadata (green), feature vector (orange) and labels (red).

by only using CGM, meal, and insulin data, it is possible to predict the quality of overnight glycaemic control.

The complete set of 19 features extracted from the day-time period was used as input to the classification problem. Then, in the first classification problem (Night_in), nights with more than 80% time in the glucose target range [70 – 180] mg/dl (time_in) with no nocturnal hypoglycaemic events detected or reported (hypo_correct, hypo event), were labelled as *On-target*. Otherwise, nights were labelled as *Off-target*. In the second classification problem (Hypo_night), nights were labelled as *Hypo* when a hypoglycaemic event was detected or reported by the user. Hypoglycaemic events were considered when there was a self-reported hypoglycaemic event or CGM measurements were below 70 mg/dL for at least 10 min, or below 55 mg/dL for at least 30 min, as defined by Maahs et al [26]. Otherwise, nights were labelled as *Non-hypo*. In the third classification problem (Hyper_night), nights were labelled as *Hyper* when the percentage time above target 180 mg/dL (time_above) during the night-time period was more than 30%. Otherwise, nights were labelled as *Non-hyper*. A description of the features and thresholds used for each type of label is reported in Table II.

In summary, the feature extraction and labeling steps produce, for each classification problem, a set of observations consisting of 19 features extracted from the day-time period and three binary labels determined from the following night-time period. Fig. 2 depicts the three previously described binary classification problems.

D. Selected Binary Classification Methods

A binary classifier is a type of supervised learning algorithm that is used to classify the elements of a given set into two groups on the basis of a classification rule [27]. In this study, three binary classifiers are used to classify the quality of overnight glycaemic control into: On-target/Off-target nights, Hypo/Non-hypo nights, and Hyper/Non-hyper nights. In

TABLE II
DEFINITION OF THE THREE CLASSIFICATION PROBLEMS FOR PREDICTING OVERNIGHT GLYCAEMIC CONTROL QUALITY

Classification problem	Abbreviation	Binary labels and definitions	
1. Night in target	Night_in	On-target: time_in > 80% & hypo_correc = 0 & no reported hypo event	Off-target: time_in ≤ 80% or hypo_correc ≠ 0 or & reported hypo event
		Hypo: hypo_correc = 0 & no reported hypo event	Non-hypo: hypo_correc ≠ 0 or reported hypo event
3. Nocturnal hyperglycaemia	Hyper_night	Hyper: time_above < 30%	Non-hyper: time_above ≥ 30%

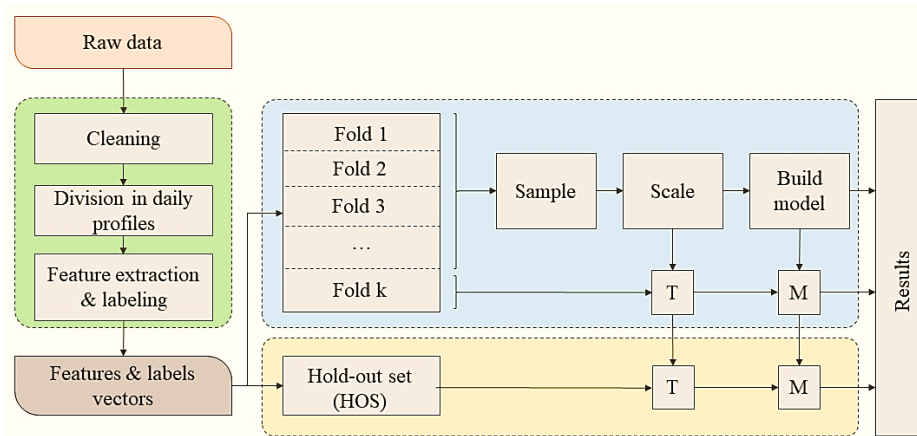


Fig. 3. Diagram of the work-flow followed in this study. First, raw-data pre-processing was performed, including removal of outliers, division in daily profiles and feature extraction. The observations are then split into 10-fold Cross-Validation Set (CVS) and Hold-out Set (HOS). Sampling and scaling were carried out after this division. The performance of the models was evaluated using completely unseen observations from HOS. T: equation to scale new observations; M: built models.

particular, the following commonly used binary classifiers have been selected for comparison purposes: Random Forest Classifier (RFC) [28], Artificial Neural Networks (ANN) [29], Support Vector Machine (SVM) [30], Linear Logistic Regression (LLR) [31], and Extended Tree Classifier (ETC) [32]. Unlike other machine learning techniques (e.g. Deep Learning), the selected techniques are well-suited to deal with relatively small datasets, such as the OhioT1DM dataset.

E. Pre-processing of the Feature Vectors

Some pre-processing techniques were applied to the feature vectors before building the models to address some issues that might arise from working with clinical datasets:

- **Class imbalance:** To handle class imbalance the Synthetic Minority Over-sampling (SMOTE) technique, which performs over-sampling of the minority class to equalize the number of samples in all the classes and acquire better performance [33].
- **Data scaling:** it is a common pre-processing step for many classifiers that can affect the classifier’s performance. Data standardization was performed on the input feature vectors. The advantage of using standard scaling is to centralize the data distribution so that the feature distribution becomes zero mean and unit variance.

F. Evaluating Performance for Model Selection

The work-flow proposed by Hernandez et al. was used to build and evaluate the models (see Fig. 3) [34]. In this process, the data was initially divided into cross-validation (CVS) and hold-out datasets (HOS), with the latter dataset comprising the 25% of the observations. It is worth noting that data-sampling was performed exclusively within the CVS training dataset. Otherwise, if applied before it would negatively affect the results by leading to over-fitting and/or generation of artificial observations that would be used for testing. In this study, 10-fold stratified cross-validation was used to assess the model’s performance on an independent dataset. Models were tested on both balanced and imbalanced versions of the testing fold to reduce the risk of over-fitting. Finally, to assess the potential of translation of the results into a clinical decision support system, the models were validated on the HOS, which contains observations that were completely unseen during the model training. The overall performance score for each model was obtained by averaging the results obtained on the HOS across the 10 folds.

G. Evaluation Metrics

A set of widely-accepted scores to evaluate the performance of binary classifiers was used. In particular, to describe and compare the model’s performance, the following metrics were employed: the area under the receiver operating characteristic

TABLE III
NUMBER OF OBSERVATIONS (DAILY PROFILES) FOR EACH SUBJECT AND WINDOW LENGTH DEFINING THE DAY-TIME PERIOD

Subject	#559	#563	#570	#575	#588	#591	Total
Raw-data	52	56	51	56	56	55	326
1h	45	49	44	46	52	45	281
3h	43	48	42	46	52	44	275
5h	42	46	41	45	52	44	270
8h	41	45	39	45	51	43	264
12h	39	43	39	42	49	39	251
18h	32	39	33	36	43	37	220

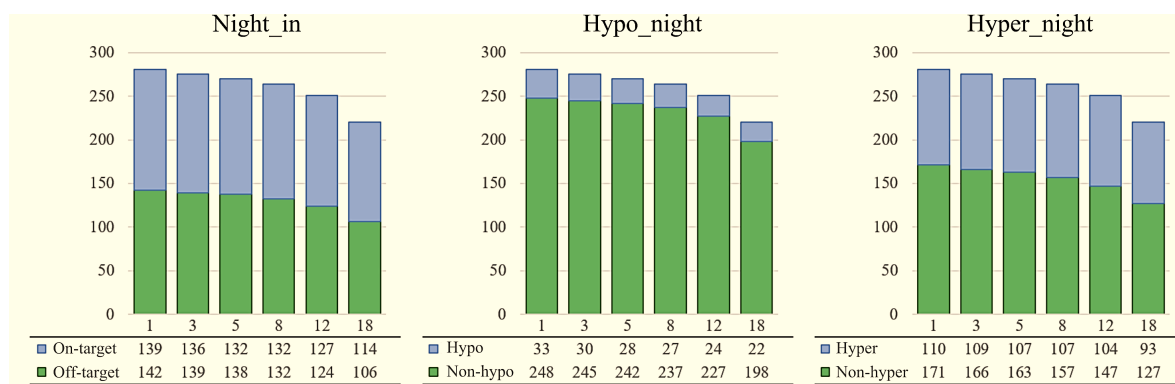


Fig. 4. Overview of class distribution within each temporal window for each label.

curve (AUC_ROC), that quantifies class separability (unitless); sensitivity (SENS), that measures the proportion of true positives that are correctly classified as such (unitless); specificity (SPEC), that measures the proportion of true negatives that are correctly classified as such (unitless); and their geometric mean (GMEAN) (unitless). It is worth highlighting that these metrics are not affected by class imbalance. In particular, the AUC_ROC enables evaluation of the classifier’s performance along its whole operating range, therefore providing greater insight into the classifier’s general performance. For each window, the model with the highest combined performance score (CS) (unitless), as defined in Equation 1, was selected as the best model.

$$CS = \sqrt{AUC_ROC^2 + GMEAN^2} \quad (1)$$

Finally, the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN), was also considered for the selection of the best models.

H. Statistical Analysis

The statistical significance of the differences between the classifiers scores was determined using the non-parametric test Kruskal-Wallis one-way ANOVA on ranks to account for the heterogeneity of variance. Post-hoc analysis using Tukey’s test was performed to determine pairwise differences. Significance level was set at $p < 0.05$.

I. Software

Python was used in this study. The implementation of the models and their evaluation was done using the models and performance scores from the library Scikit-learn [35]. Moreover, the sampling techniques were used from the library Imbalanced-learn [36], data handling was done with Pandas [37] and data visualization using Matplotlib and Seaborn [38]. Finally, the statistical analysis was performed using Statsmodels [39].

III. RESULTS

A. Data Insight

The number of observations (i.e. daily profiles) before and after the raw-data pre-processing (i.e., after profiles with too many missing data were discarded) for each of the subjects in the dataset (#559, #563, #570, #575 #588, #591) and for the different time windows lengths defining the day-time (1h, 3h,

5h, 8h, 12h, 18h) is shown in Table III. The number of on-target and off-target nights for each day-time window and label is depicted in Fig. 4.

B. Prediction of Quality of Overnight Glycaemic Control

Table 4 presents the model with the best performance based on the combined score (CS) of AUC_ROC and GMEAN (Equation 1) for each classification problem (Night_in, Hypo_night, Hyper_night) and temporal window. In addition, Fig. 5 depicts the scores (AUC_ROC, GMEAN, SPEC and SENS) associated with these classifiers for the different windows lengths and labels.

As shown in Fig. 5, the best AUC_ROC score for the first two classification problems (glucose in target – Night_in, and presence of nocturnal hypoglycaemia – Hypo_night) is achieved with a window of 18h using the ETC and SVM classifiers respectively. These classifiers also present the highest GMEAN values (around 0.65 in both cases), but the results are not significantly different to the performance in other windows. Regarding the classification of night in target (Night_in), both specificity and sensitivity are moderate (between 0.5 and 0.65). On the other hand, for the classification of nocturnal hypoglycaemia (Hypo_night), the specificity is greater than the sensitivity for every window length. Focusing on the results of the third classification problem for predicting nocturnal hyperglycaemia (Hyper_night), the best model was found to be the RFC classifier in the window of 8h. While the sensitivity was lower for all the window lengths (0.5 to 0.6), the AUC_ROC (0.73) and the specificity (0.8) obtained with this classifier were significantly higher than those obtained by the other classifiers.

IV. DISCUSSION

Pre-processing of the data used for training and evaluating the binary classifiers, resulted in a considerable reduction of the number of observations when compared to the original dataset. This is an indicator of the significant number of missing CGM data points. Moreover, there is also a decrease in the number of observations with the increase of the day-time window length. This outcome is expected, because the longer the considered time interval, the higher the probability of having two-hour CGM gaps. Moreover, using splines to interpolate missing CGM data might have introduced variability that can affect the classification performance. Future work could take advantage of more advanced interpolation techniques, such as the

TABLE IV
OVERVIEW OF BEST CLASSIFIERS (SHADOWED AREA) FOR EACH TEMPORAL WINDOW AND CLASSIFICATION PROBLEM

Problem	Window	Best model	Configuration
Night_in	1h	RFC	Minimum number of samples per leaf: 1; Number of trees: 50
	3h	RFC	Minimum number of samples per leaf: 10; Number of trees: 20
	5h	ANN	Activation: logistic sigmoid function; Alpha=0.0001; Learning rate: 0.001; Solver: stochastic gradient-based optimizer
	8h	ETC	Minimum number of samples per leaf: 50; Number of trees: 50
	12h	ETC	Minimum number of samples per leaf: 5; Number of trees: 50
	18h	ETC	Minimum number of samples per leaf: 50; Number of trees: 50
Hypo_night	1h	ANN	Activation: logistic sigmoid function; Alpha=0.1; Learning rate: 0.001; Hidden layer sizes: (10, 10); Solver: stochastic gradient-based optimizer
	3h	ANN	Activation: logistic sigmoid function Alpha=0.1; Learning rate: 0.001; Hidden layer sizes: (5,0); Solver: stochastic gradient-based optimizer
	5h	SVM	Decision function shape: one vs rest; Kernel coefficient: 0.01, Kernel: Radial-basis function
	8h	RFC	Minimum number of samples per leaf: 5; Number of trees: 10
	12h	LLR	Optimization algorithm: Newton-CG
Hyper_night	1h	ETC	Minimum number of samples per leaf: 5; Number of trees: 10
	3h	ETC	Minimum number of samples per leaf: 50; Number of trees: 10
	5h	ETC	Minimum number of samples per leaf: 5; Number of trees: 50
	8h	RFC	Minimum number of samples per leaf: 10; Number of trees: 50
	12h	ETC	Minimum number of samples per leaf: 50; Number of trees: 100
	18h	ETC	Minimum number of samples per leaf: 50; Number of trees: 100

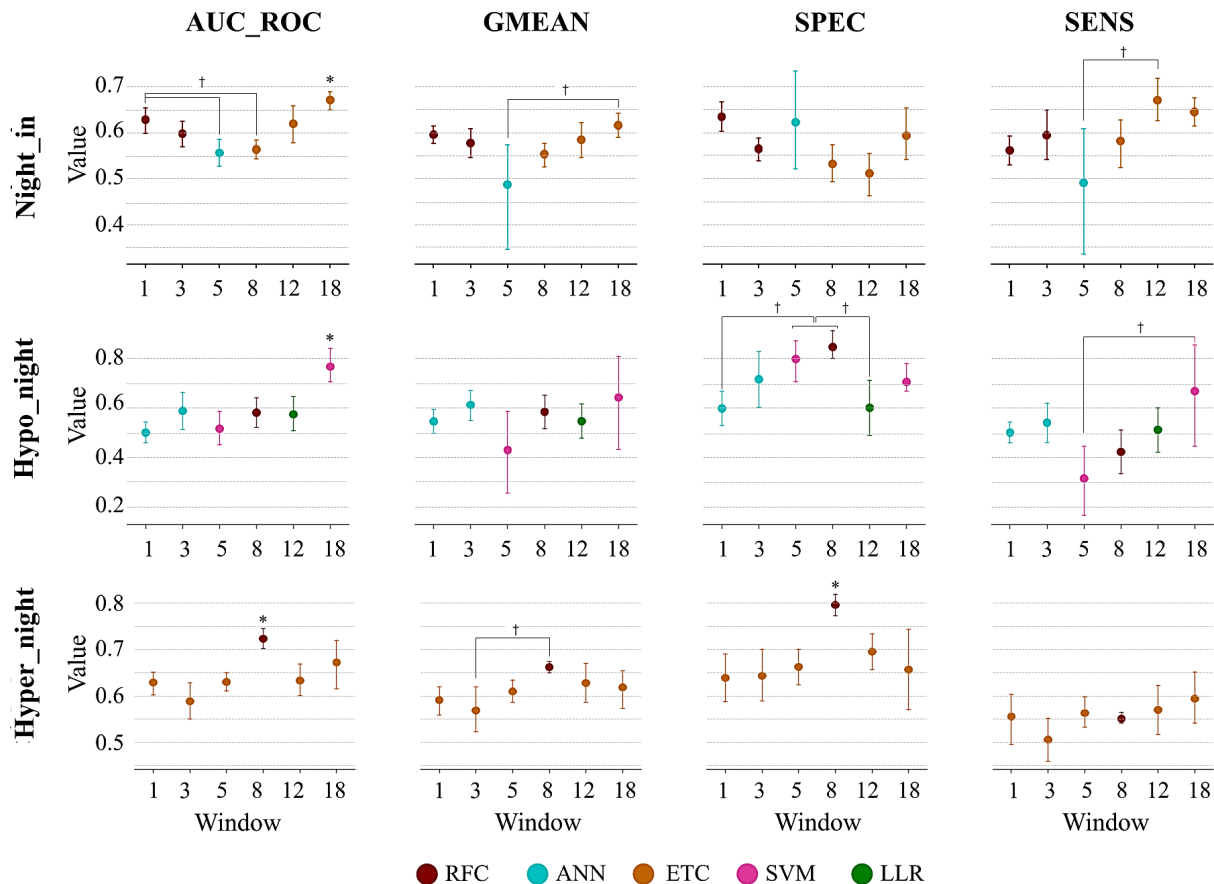


Fig. 5. Scores obtained with the best classifier for each temporal window length and label. *: Differences were significant in all the post-hoc pair comparisons ($p < 0.05$). †: Differences were significant only between the linked classifiers ($p < 0.05$).

use of T1DM physiological models to account for the glucose dynamics.

It is interesting to note that in the first classification problem (Night_in), there is a balanced distribution of the labels (i.e. normoglycaemia vs. abnormal glycaemia) along all the windows (see Fig. 4). However, some imbalance in the distributions arises in the second classification problem based on presence of a nocturnal hyperglycaemic event (around 40% of nights are hyperglycaemic). This imbalance is substantial when classifying based on the presence of hypoglycaemic nights (only around 10% of nights have hypoglycaemic events).

The impact of the window length in each classification problem was extensively studied. As depicted in Fig. 5, better results are obtained with larger windows, which might be explained by the fact that they have more information from the events occurred throughout the day. However, note that the larger the window, the smaller the dataset used for training and evaluating the models. Hence, the impact of the size of the dataset on the results is something that should be further evaluated in future studies.

In terms of model selection, as reported in Table 4, the best model for each window length greatly varies for each classification problem (Night_in, Hypo_night, Hyper_night). As an example, for window lengths of 1h and 3h the best models are RFC, ANN and ETC for each problem respectively. Moreover, within each classification problem there is also a great variation on the best model across window lengths, except when classifying based on nocturnal hyperglycaemia. In this case, the best model for almost all windows, excluding the one of 8h, was found to be an ETC with different parameters' configuration. For the 8h window the best model was an RFC classifier, presenting significantly higher AUC_ROC and specificity in comparison with the best models on other windows.

Interestingly, for the classification problems based on the percentage of time in target and the presence of nocturnal hypoglycaemia, the highest AUC_ROC and GMEAN scores are achieved in the window of 18h using an ETC and SVM classifiers, respectively. In these classification problems, the sensitivity achieved with the best model for each window ranges from 0.3 up to almost 0.7, suggesting great dependence in the dataset used. Finally, the obtained specificity was overall higher, especially for classification based on nocturnal hypo and hyperglycaemia, where values up to 0.8 are obtained.

Despite the promising results suggesting that different aspects of the quality of overnight glycaemia can be predicted, there are some limitations that need to be addressed before translating this methodology into real practice. The main drawback is associated with the scarce amount of available data, which greatly restricts the training and validation of the classifiers. As an example, for the first classification problem (Night_in), in the testing set for the window of 18 hours there were only 3 observations with the hypoglycaemia label in a total of 58 observations. It is very difficult for the algorithms to infer patterns for further classification from such a small dataset. Other methods that could be investigated are the development of multi-class and ensembled classifiers, which could potentially improve the results. Another way to improve the results is the inclusion of additional features. For example, incorporating information about changes in the basal insulin

profile could help improving the classification of overnight glycaemic control. Moreover, non-stationary frequency analysis methods, such as wavelets transform, which provide temporal-specific glycaemic features, could also potentially improve the classification results. Such potential improvement will be investigated in future work. The OhioT1DM dataset includes other physiological data, such as exercise and stress, which could also be included to improve the prediction of overnight glycaemic control. However, initial studies carried out by us, and by other authors [40], indicate that the inclusion of these parameters do not yield better results. Finally, if reliably measured, the quality of sleep over the preceding nights could also be accounted as a predictor for the glucose control of the current night.

In terms of practical implementation of the proposed technique in a real-life setting, this could be embedded within a mobile app, or within a sensor-augmented insulin pump, that interfaces to a CGM device. Then, by indicating bedtime to the app, the user would get back information about the probability of having a night within target, having hypoglycaemia, and having hyperglycaemia. In addition, based on these probabilities, a simple heuristic could be implemented to recommend the required actions to the user in order to have a smooth night (e.g., reducing the basal insulin by 30%).

V. CONCLUSION

This study indicates that it is possible to predict the quality of overnight glycaemia by using data from the continuous glucose measurements (CGM) and user-reported data (carbohydrates and insulin) from the preceding day-time period. Although there is no clearly superior approach among the evaluated machine learning techniques (binary classifiers), Extended Tree Classifier (ETC) and Support Vector Machine (SVM) perform better at predicting normoglycaemia and hypoglycaemia, while Random Forest Classifier (RFC) performs better at predicting nocturnal hyperglycaemia. Regarding the length of the considered day-time window, the best results were obtained for longer windows (e.g. 18 hours). Despite the promising results, a larger dataset is needed in order to fully validate the proposed approach. Finally, with a larger dataset, more advanced classification algorithms suited for processing sequences of data, such as long short-term memory networks (LSTM), can be evaluated.

REFERENCES

- [1] M. A. Atkinson et al., "Type 1 diabetes," *Lancet*, vol. 383, no. 9911, pp. 69-82, Jan. 2014, DOI: 10.1016/S0140-6736(13)60591-7 [Online].
- [2] M. Vettoretti et al., "Continuous glucose monitoring: Current use in diabetes management and possible future applications," *J Diabetes Sci Technol*, vol. 12, no. 5, pp. 1064-1071, Sep. 2018, DOI: 10.1177/1932296818774078 [Online].
- [3] G. P. Forlenza et al., "Predictive low-glucose suspend reduces hypoglycemia in adults, adolescents, and children with type 1 diabetes in an at-home randomized crossover study: Results of the PROLOG trial," *Diabetes Care*, vol. 41, no. 10, pp. 2155-2161, Oct. 2018, DOI: 10.2337/dc18-0771 [Online].
- [4] B. Kovatchev et al., "Automated closed-loop control of diabetes: The artificial pancreas," *Bioelectronic Medicine*, vol. 4, no. 14, pp. 1-12, Nov. 2018, DOI: 10.1186/s42234-018-0015-6 [Online].
- [5] D. C. Kovatchev, "The current status of bolus calculator decision-support software," *J Diabetes Sci Technol*, vol. 6, no. 5, pp. 990-994, Sep. 2012, DOI: 10.1177/193229681200600501 [Online].

- [6] D. Lewis, "Setting expectations for successful artificial pancreas/hybrid closed loop/automated insulin delivery adoption," *J Diabetes Sci Technol*, vol. 12, no. 2, pp. 533-534, Mar. 2018, DOI: 10.1177/1932296817730083 [Online].
- [7] I. Contreras et al., "Artificial intelligence for diabetes management and decision support: Literature review," *J Med Internet Res*, vol. 20, no. 5, pp. e10775, May 2018, DOI: 10.2196/10775 [Online].
- [8] G. Cappon et al., "Wearable continuous glucose monitoring sensors: A revolution in diabetes treatment," *Electronics*, vol. 6, no. 3, pp. 65, Sep. 2017, DOI: 10.3390/electronics6030065, [Online].
- [9] P. Herrero et al., "Advanced insulin bolus advisor based on run-to-run control and case-based reasoning," *IEEE J Biomed Health Inform*, vol. 19, no. 3, pp. 1087-1096, May 2015, DOI: 10.1109/JBHI.2014.2331896 [Online].
- [10] G. Cappon et al., "A neural-network-based approach to personalize insulin bolus calculation using continuous glucose monitoring," *J Diabetes Sci Technol*, vol. 12, no. 2, pp. 265-272, Mar. 2018, DOI: 10.1177/1932296818759558 [Online].
- [11] R. A. Lal et al., "Advances in care for insulin-requiring patients without closed loop," *Diabetes Technol Ther.*, vol. 20, no. S2, pp. S285-S291, Jun. 2018, DOI: 10.1089/dia.2018.0084 [Online].
- [12] K. Turksoy et al., "Hypoglycemia detection and carbohydrate suggestion in an artificial pancreas," *J Diabetes Sci Technol*, vol. 10, no. 6, pp. 1236-1244, Nov. 2016, DOI: 10.1177/1932296816658666 [Online].
- [13] A. Beneyto et al., "A new blood glucose control scheme for unannounced exercise in type 1 diabetic subjects," *IEEE Trans Control Syst Technol* pp. 1-8, Nov. 2018, DOI: 10.1109/TCST.2018.2878205 [Online].
- [14] C. Liu et al., "A modular safety system for an insulin dose recommender: a feasibility study," *J Diabetes Sci Technol*, 2019, (in press).
- [15] M. D. Breton et al., "Continuous glucose monitoring and insulin informed advisory system with automated titration and dosing of insulin reduces glucose variability in type 1 diabetes mellitus," *Diabetes Technol Ther*, vol. 20, no. 8, pp. 531-540, Aug. 2018, DOI: 10.1089/dia.2018.0079 [Online].
- [16] JDRF CGM Study Group, "Prolonged nocturnal hypoglycemia is common during 12 months of continuous glucose monitoring in children and adults with type 1 diabetes," *Diabetes Care*, vol. 33, no. 5, pp. 1004-1008, DOI: 10.2337/dc09-2081 [Online].
- [17] E. Tsalikian et al., "Impact of exercise on overnight glycemic control in children with type 1 diabetes mellitus," *J Pediatr*, vol. 147, no. 4, pp. 528-534, Oct. 2005, DOI: 10.1016/j.jpeds.2005.04.065 [Online].
- [18] S. Oviedo et al., "A review of personalized blood glucose prediction strategies for T1DM patients," *Int J Numer Method Biomed Eng*, vol. 33, no. 6, Jun. 2017, DOI: 10.1002/cnm.2833 [Online].
- [19] K. Li et al. "Convolutional recurrent neural networks for glucose prediction," *IEEE J Biomed Health Inform*, Apr. 2019, DOI: 10.1109/JBHI.2019.2908488 [Online].
- [20] S. R. Newcomer et al., "Identifying subgroups of complex patients with cluster analysis," *Am J Manag Care*, vol. 17, no. 8, pp. e324-e332, Aug. 2011.
- [21] L. Meneghetti et al., "Data-driven anomaly recognition for unsupervised model-free fault detection in artificial pancreas," *IEEE Trans Control Syst Technol*, pp. 1-15, Dec. 2018, DOI: 10.1109/TCST.2018.2885963 [Online].
- [22] A. Bertachi et al., "Prediction of blood glucose levels and nocturnal hypoglycemia using physiological models and artificial neural networks," *The 3rd International Workshop on Knowledge Discovery in Healthcare Data*, Stockholm, Sweden, July 13, 2018.
- [23] C. Marling et al., "The OhioT1DM dataset for blood glucose level prediction," *The 3rd International Workshop on Knowledge Discovery in Healthcare Data*, Stockholm, Sweden, July 13, 2018.
- [24] R. Hovorka et al., "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiol Meas*, vol. 25, no. 4, pp. 905-920, Aug. 2004.
- [25] G. Fico, L. Hernandez, J. Cancela, M. Isabel, A. Facchinetti, C. Fabris, R. Gabriel, C. Cobelli and M. Arredondo. Exploring the Frequency Domain of Continuous Glucose Monitoring Signals to Improve Characterization of Glucose Variability and of Diabetic Profiles *J. Diabetes Sci. Technol*. Vol 0, 2017
- [26] D. M. Maahs et al., "Outcome measures for artificial pancreas clinical trials: a consensus report," *Diabetes Care*, vol. 39, no. 7, pp. 1175-1179, Jul. 2016, DOI: 10.2337/dc15-2716 [Online].
- [27] T. Hastie et al., "The element of statistical learning: Data mining, inference, and prediction," *Springer Series in Statistics*, 2001, <https://web.stanford.edu/~hastie/Papers/ESLII.pdf> [Online].
- [28] L. Breiman. Random forests. *Machine Learning*, 45:5-32, 2001.
- [29] Haykin, Simon. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.
- [30] A. J. Smola, and B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3), 199-222, 2004
- [31] R. B. Darlington. *Regression and linear models*. New York: McGraw-Hill., pp. 292-293, 1990
- [32] L. Breiman. Random forests. *Machine Learning*, 45:5-32, 2001.
- [33] N. V. Chawla et al., "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, Jun. 2002, DOI: 10.1613/jair.953 [Online].
- [34] B. Hernandez et al., "Supervised learning for infection risk inference using pathology data," *BMC Med Inform Decis Mak*, vol. 17, no. 1, pp. 168, Dec. 2017, DOI: 10.1186/s12911-017-0550-1 [Online].
- [35] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, Oct. 2011.
- [36] G. Lemaitre et al., "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, pp. 1-5, Jan. 2017.
- [37] W. McKinney et al., "Pandas: a foundational Python library for data analysis and statistics," *Python for high performance and scientific computing*, 2011.
- [38] M. D. Bloice et al., "A tutorial on machine learning and data science tools with python," in *Machine Learning for Health Informatics*, Springer, pp. 435-480, Dec. 2016.
- [39] S. Seabold et al., "Statsmodels: Econometric and statistical modeling with python," in *Proceedings of the 9th Python in Science Conference*, vol. 57, pp. 61, Jan. 2010.
- [40] C. Midroni, P. J. Leimbigler, G. Baruah, M. Kolla, A. J. Whitehead, & Y. Fossat. Predicting Glycemia in Type 1 Diabetes Patients: Experiments with XGBoost. *KDH@IJCAI-ECAI 2018 The 3rd International Workshop on Knowledge Discovery in Healthcare Data*, Vol. 2148. Stockholm, Sweden, July 13, 2018.