Prediction of Nocturnal Hypoglycaemia in Adults with Type 1 Diabetes using Machine Learning Classifiers

Ioannis Afentakis, Rebecca Unsworth, Pau Herrero, Nick Oliver, Monica Reddy, Pantelis Georgiou

***Abstract*— *Objective:* One of the biggest challenges for people with Type 1 Diabetes (T1D) using multiple daily injection (MDI) therapy is nocturnal hypoglycaemia (NH). Recurrent nocturnal hypoglycaemia can lead to serious complications and so prevention is of high importance. This work aims to provide bedtime decision support to people with T1D, to minimize the risk of NH. *Methods:* We present the design and development of binary classifiers that can be used to predict NH (blood glucose levels occurring below 3.9 mmol/L). Using data collected from a 6-months study of adult participants with T1D under free-living conditions, we extract daytime features from continuous glucose monitor (CGM) sensors, administered insulin, meal and physical activity information. We use these features to train and test two machine learning algorithms; Random Forests (RF) and Support Vector Machines (SVM). *Results:* At population-level model, SVM outperforms RF algorithm with a ROC-AUC 79.36% (95% CI, 76.86% - 81.86%). We further evaluate our model in a different population of 20 adults with T1D using MDI insulin therapy and wearing CGM and flash glucose monitoring sensors for two periods of 8 weeks each. *Conclusion:* Our model shows state-of-the-art performance and generalizability in a completely unseen dataset, and is robust when tested in sensor devices from different manufacturers. *Significance:* The proposed algorithm is a potential viable approach to inform people with T1D about their risk of NH before it occurs.**

***Index Terms*— machine learning, nocturnal hypoglycaemia, Type 1 diabetes**

I. Introduction

People with type 1 diabetes (T1D) rely on exogenous insulin therapy, aiming to maintain blood glucose concentrations within a target range. There are various advanced technologies for insulin treatment including continuous glucose sensors, continuous subcutaneous insulin infusion pumps (CSII), and closed-loop systems (i.e., artificial pancreas). Most people intermittently inject insulin in a multiple daily injection (MDI) regimen. One of the biggest challenges of MDI therapy is hypoglycaemia [1] (usually defined when blood glucose levels fall below 3.9 mmol/L). Most hypoglycaemic episodes in people with T1D occur during sleep [2] even though there are advanced sensing technologies able to detect these events. Nocturnal hypoglycaemia is a great challenge for people with T1D [3], since during sleep hypoglycaemia awareness is attenuated. If left untreated and prolonged, recurrent exposure to hypoglycaemia is associate with impaired awareness of hypoglycaemia and the dead-in-bed syndrome.

The extensive use of continuous glucose monitors (CGM) has led to a plethora of large and rich datasets which can be exploited to train supervised machine learning algorithms to predict adverse events before they occur. Previous studies have focused mainly on the prediction of blood glucose levels in a short-term prediction horizon. However, there are only a few research studies proposing models for predicting nocturnal hypoglycaemia with a longer horizon (e.g., during sleep). Due to the high inter-subject and intra-subject variability, the number of subjects in a dataset as well as the number of observations per subject (i.e., day-night profiles) play a crucial role in the development of algorithms.

Vu L. *et al*, used a large dataset of 10,000 CGM users with over a million nights, and built a Random Forest binary classifier to predict the occurrence of NH [4]. The proposed model achieved a high ROC-AUC (Receiver Operating Characteristic Area Under the Curve) score of 84%.

A similar study by Jensen M. *et al*, used data from a clinical trial consisted of 463 participants with T1D to predict level 2 nocturnal hypoglycaemia ($\leq$ 3 mmol/L) [5]. They extracted features from the daytime as well as the three consecutive days before the night to train a linear discriminant function (LDA) algorithm. The proposed algorithm consists of four features and can be used to make predictions at midnight achieving a 79% ROC-AUC score.

A. Bertachi *et al*, have developed individualized prediction models able to detect more than 70% of nocturnal hypoglycaemia events in people with T1D [6]. They trained their algorithms on a dataset of 10 participants in a clinical study for 12 weeks under free living conditions. In total they used 29 features extracted from CGM signals, insulin, meal intake and an activity tracker, and they showed that the SVM algorithm outperforms the multilayer perceptron in that population.

A. Güemes *et al*, proposed an approach for predicting the quality of overnight glycaemic control in people with T1D using binary classifiers [7]. Using the publicly available clinical dataset (OhioT1DM) [8], they extracted features from CGM measurements, insulin dosage and carbohydrate intake and trained multiple machine learning models aiming to predict the presence of nocturnal hypoglycaemia or hypergycemia as well as the percentage time in range (%TIR). The authors did not find any strict superiority between the algorithms they tested, and they report an overall ROC-AUC score of around 70%.

C. Mosquera-Lopez *et al*, used a large dataset collected from 124 people with T1D, consisted around 23 thousand nights, and used glucose, insulin, and meal information to train an SVR model [9]. The output of the model was optimized to maximize the benefit of an accurate nocturnal hypoglycaemia prediction and to minimize the cost of an inaccurately predicted event using decision theory. They tested their model in-silico, and the results showed that the proposed algorithm can reduce 77% of nocturnal hypoglycaemia events without impacting time in range (TIR).

In this work we propose an algorithm, to provide bedtime decision support to people with T1D and minimize the risk of NH. We use a novel open-source framework for extracting features from blood glucose time series, which can be used for the prediction of NH before bedtime as well as other adverse events. We also show that the algorithm is robust when tested in a dataset of a different population and we prove the ability of the model to be used with sensors from different manufacturers.

Manuscript received March, 2023. “This work was supported by the UKRI Center for Doctoral Training in AI for Healthcare http://ai4health.io Grant number EP/S023283/1”(Corresponding author: I. Afentakis)

The author I. Afentakis is with the UKRI CDT in AI for Healthcare and the Department of Computing, Imperial College London, London SW7 2RH, U.K. (e-mail: i.afentakis20@imperial.ac.uk).

The authors P. Herrero and P. Georgiou are with the Department of Electronic and Electrical Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: p.herrero-vinias@imperial.ac.uk; pantelis@imperial.ac.uk)

The authors R. Unsworth, N. Oliver, M. Reddy are with the Department of Medicine, Imperial College London, London SW7 2BX, U.K. (e-mail: r.unsworth@imperial.ac.uk; nick.oliver@imperial.ac.uk; m.reddy@imperial.ac.uk)

II. Materials and Methods

A. Datasets

The main dataset used for the development of the algorithms has been collected from a 6-months randomized controlled, cross-over study (Clinical Trial Registry No: NCT03963219, ethics approval from the regional ethics committee and Medicine and Healthcare products Regulatory Agency (MHRA)), in participants wearing continuous glucose monitoring (CGM) sensors and using an intensified MDI regimen. A total of 37 adult participants used Dexcom G6 CGM sensors (Dexcom Inc, San Diego, CA, US) as well as a Fitbit Charge 3 (Fitbit, Inc., San Francisco, CA, USA) activity tracker. Additionally, they used a smartphone app where to record administered insulin and meal macronutrient information. Participant demographics and clinical data are shown in Table I.

Participants enrolled in the study also met the following inclusion criteria; diagnosed with T1D for more than 3 years, undertaken structured education and on intensified multiple dose insulin injection regimen for more than 6 months. Exclusion criteria included pregnancy or breastfeeding, a history of renal impairment, uncontrolled thyroid disease, or Ischaemic heart disease, as well as an allergy or intolerance to insulin aspart. Participants with visual impairment, active malignancy, reduced manual dexterity, or those enrolled in other clinical trials were not considered.

To further validate our model, provide evidence about its generalization power and to test its performance in different glucose monitoring sensor devices, we used a dataset collected from a head-to-head glucose monitoring study (Clinical Trial Registry No: NCT03028220, ethics approval from London - Hampstead Research Ethics Committee (Reference no.: 15/LO/1679)), comparing CGM and flash glucose monitoring [10], [11]. In this study 40 high-risk adult participants using MDI for their insulin treatment, were randomly assigned to CGM Dexcom G5 (Dexcom Inc, San Diego, CA, US) or FreeStyle Libre 1 (Abbott Diabetes Care, Alameda, CA, USA) intermittently scanned CGM for 8 weeks. An extension phase was conducted after finishing the main study, during which all participants used Dexcom G5 for an additional 8 weeks. A description of this testing dataset is presented in Table II.

TABLE I
PARTICIPANT DEMOGRAPHICS OF MAIN DEVELOPMENT DATASET

Data	Values
Demographics	
Age (mean)	36 (29-46)
Gender (female/male)	15/22 (40% female)
BMI	26.63 $\pm$ 5.18
Clinical	
HbA1c (mmol/mol)	61 (52-66)
eA1c (%)	7.14 $\pm$ 0.87
T1D duration (years)	$>$ 3
Time below range (%)	4.97 $\pm$ 3.93
Time in range (%)	63.15 $\pm$ 15.50
Time above range (%)	31.88 $\pm$ 16.02
Low blood glucose index	1.34 $\pm$ 0.89
High blood glucose index	7.37 $\pm$ 3.94

TABLE II
PARTICIPANT DEMOGRAPHICS OF FURTHER VALIDATION DATASET

Data	Values
Demographics	
Age (mean)	49 (37-63)
Gender (female/male)	16/24 (40% female)
Clinical	
HbA1c (mmol/mol)	56 (48-63)
eA1c (%)	7.3 $\pm$ 0.50
T1D duration (years)	$>$ 20

B. Data pre-processing

The continuous glucose data were collected and exported as time series, resulting in 288 glucose readings per day, one every five minutes. However, due to some sensors malfunctions or replacements (for example during the first two hours after inserting a new sensor there is no signal) some days contain missing values. To address this issue linear interpolation was used to impute missing data for up to 6 consecutive samples (i.e., 30 minutes). Cases including longer periods of missing values were excluded from the dataset (accounted for less than 5%).

Outliers in the participants' self-reported insulin values or consumed carbohydrates were identified based on the interquartile rule (observations 1.5 times the IQR below or above the first and third quartile respectively). These values were consequently imputed with the mean or median of each participant's data. Additionally, days when participants consumed a meal reported as snack after their dinner, which was not followed by an insulin injection were also excluded from the dataset. We assumed these were cases of rescue carbohydrates used as an intervention from participants due to fear of NH. A total of 418 nights fell in this category and thus excluded from the training set.

In contrast with the CGM sensors, flash sensors do not sync data automatically with a reader or a mobile app. A scan once every 8 hours is required to get the complete glycaemic picture of the day. Also, their sampling frequency is 15 minutes instead of 5 minutes for CGM sensors. In order to use the second testing dataset to evaluate our models a couple of adaptations were performed in the data coming from the flash sensors. First, we disregarded all the incomplete daily profiles of users who did not perform the required number of daily scans (i.e., 3-4 scans) and resulted in hours of missing values. We also resampled the flash signal with a 5-minute frequency by interpolating the intermediate values, so that flash time series follow the CGM time series format (i.e., 288 readings per day).

The labelling of the target class was performed as follows. For every daily profile of each participant, the period between midnight and 6am was considered to be night-time. During this night-time if there was at least one period of 20 consecutive minutes (or more), with glucose levels falling below 3.9 mmol/L, then this night was labelled as a hypoglycaemic night (Class 1), otherwise as a night with absence of any hypoglycaemic episode (Class 0). The remaining daytime hours from 6am to midnight were used for features extraction.

C. Feature Extraction

1) Glucose Features: Glucose levels collected from the Dexcom G6 CGM sensor were used to extract glucose-related features for each participant. For every daytime period, defined from 6am to midnight, a set of time series features across the temporal, statistical and spectral domain were extracted using a novel feature engineering machine learning framework [12]. Additionally, commonly used diabetes-related metrics were considered, such as percentage Time in Range (TIR), percentage Time below Range (TBR) and Time above Range (TAR), (i.e., the percentage time where glucose levels fall within, below, or above the target range [3.9, 10] mmol/L) as well as high and low blood glucose indexes [13] among others. The full list of features is shown in Table III. To better capture the full signal from the CGM and the glucose information related to diabetes, an iterative and recursive approach was followed. Specifically, all the time series and diabetes-related features were calculated for every single time window starting from the last hour prior to sleep, the

last 2 hours prior to sleep, up to the last 12 hours prior to sleep.

2) Insulin Features: Information about the insulin doses that participants inject was collected for the development dataset. In the app, users are responsible for self-reporting the units of insulin they inject based on the recommendations they receive from the standard bolus calculator or the Advanced Bolus Calculator for Diabetes. To extract insulin-related features, an approximation of the insulin on board (IOB) model [14] was used. Specifically, considering the four-hour time window before sleep (8pm - midnight) the IOB at midnight was calculated, assuming a linear decay of four hours from the time of injection. In case of multiple injections, the summary of the insulin was computed. Basal insulin levels were kept unchanged during the study for all participants, and so we did not include them in the features set.

3) Meal Intake Features: Similarly, with the insulin boluses, meal intake information is also reported manually by participants in the mobile app. After each meal users have been requested to log the amount of carbohydrates they consumed. Carbohydrates on board (COB) were calculated for every participant as the total amount of carbohydrates (summary of carbohydrate in grams) they consumed between 6pm and midnight. This time window was selected after observing that in the majority of daily profiles, participants have their last meal of the day between this time frame.

4) Physical Activity Features: Participants of the study have also been requested to wear a Fitbit Charge 3 (Fitbit, Inc., San Francisco, CA, USA) activity monitor. The data were downloaded from Fitbit database and included various information about users' steps, heart rate and energy expenditure among others. For each participant, a set of features was extracted representing their daily activity as shown in Table IV.

D. Algorithm Development

A Sequential Forward Selection algorithm was used for the identification of the subset of the most relevant features for this specific learning problem. During the first iteration, every feature is tested individually in the classification task and the one with the highest ROC-AUC score is selected. At every other iteration the algorithm chooses the feature which maximizes the underlined objective (to maximize the ROC-AUC), based on the cross-validation score, and includes this feature in the previous selected subset. The resulting subset of features is shown in Table V.

The main dataset that used for the development of the algorithms consists of 37 participants, around six thousand nights in total, in 11% of which at least one hypoglycaemic episode occurred (minimum of 20 consecutive minutes where glucose levels were below 3.9 mmol/L). Hence, the distribution of the target class is unequal, as there is a 9 to 1 ratio between non hypoglycaemic nights (majority class) and nights with NH events (minority class). An imbalanced dataset can degrade the performance of the classifiers if not treated appropriately. For this reason, the Synthetic Minority Over-Sampling Technique (SMOTE) [15] was employed to generate synthetic samples for the minority class with the help of interpolation between the positive instances that lie together.

Rescaling of the range of features was also performed during the training process. Specifically, the min-max normalization formula was used as shown in (1). This step is particularly important for distance-based algorithms such as Support Vector Machines (SVM) that we use in this application. This is because such algorithms use distances between data points to determine their similarity and perform the

TABLE III
FULL LIST OF GLUCOSE-RELATED FEATURES

Feature name	Description
Time Series Features	
Abs_energy	Absolute energy of the signal
Auc	Area under the curve of the signal (trapezoid rule)
autocorr	Autocorrelation of the signal
Calc_centroid	Centroid along the time axis
Calc_max	Maximum value of the signal
Calc_mean	Mean value of the signal
Calc_median	Median value of the signal
Calc_min	Minimum value of the signal
Calc_std	Standard deviation of the signal
Calc_var	Variance of the signal
distance	Signal traveled distance
ecdf	Values of ECDF along time axis
Ecdf_percentile	Percentile values of ECDF
Ecdf_percentile_count	Cumulative sum of samples that are less than the percentile
entropy	Entropy of the signal using the Shannon Entropy
Fft_mean_coeff	Mean values of each spectrogram frequency
Fundamental_frequency	Fundamental frequency of the signal
hist	Histogram of the signal
Human_range_energy	Human range energy ratio
Interq_range	Interquartile range of the signal
kurtosis	Kurtosis of the signal
lpcc	Linear prediction cepstral coefficients
Max_frequency	Maximum frequency of the signal
Max_power_spectrum	Maximum power spectrum density of the signal
Mean_abs_deviation	Mean absolute deviation of the signal
Mean_Abs_diff	Mean absolute difference of the signal
Mean_dif	Mean of differences of the signal
Median_abs_deviation	Median absolute deviation of the signal
Median_abs_diff	Median absolute differences of the signal
Median_diff	Median differences of the signal
Median_frequency	Median frequency of the signal
mfcc	MEL cepstral coefficients
Negative_turning	Number of negative turning points of the signal
Neighbourhood_peaks	Number of peaks from a defined neighborhood of the signal
Pk_pk_distance	Peak to peak distance
Positive_turning	Number of positive turning points of the signal
Power_bandwidth	Power spectrum density bandwidth of the signal
rms	Root mean square of the signal
skewness	Skewness of the signal
slope	Slope of the signal
Spectral_decrease	Amount of decreasing of the spectral amplitude
Spectral_distance	Single spectral distance
Spectral_entropy	Spectral entropy of the signal based on Fourier transform
Spectral_kurtosis	Flatness of a distribution around its mean value
Spectral_positive_turning	Number of positive turning points of the fft magnitude signal
Spectral_roll_off	Spectral roll-off of the signal
Spectral_roll_on	Spectral roll-on of the signal
Spectral_skewness	Asymmetry of a distribution around its mean value
Spectral_slope	Spectral slope
Spectral_spread	Spread of the spectrum around its mean
Spectral_variation	Amount of variation of the spectrum along time
Sum_abs_diff	Sum of absolute differences of the signal
Total_energy	Total energy of the signal
Wavelet_abs mean	CWT absolute mean value of each wavelet scale
Wavelet_energy	CWT energy of each wavelet scale
Wavelet_entropy	CWT entropy of the signal
Wavelet_std	CWT standard deviation value of each wavelet scale
Wavelet_var	CWT variance value of each wavelet scale
Zero_cross	Zero-crossing rate of the signal
Diabetes Features	
ri	Risk index
lbgi	Low blood glucose index
hbgi	High blood glucose index
TIR	% Time in range [3.9, 10] mmol/L
Tbr_1	% time in [3, 3.9) mmol/L
Tbr_2	% time below 3 mmol/L
Tar_1	% time in (10, 13.9] mmol/L
Tar_2	% time above 13.9 mmol/L

TABLE IV
DESCRIPTION OF THE DAILY PHYSICAL ACTIVITY FEATURES

Feature	Metric
Total steps	total count
Total distance	meters
Very active distance	meters
Moderately active distance	meters
Light active distance	meters
Sedentary active distance	meters
Very active minutes	minutes
Fairly active minutes	minutes
Lightly active minutes	minutes
Sedentary active minutes	minutes
Calories - Total estimated energy expenditure	kilocalories
floors	total count
Calories BMR - Total energy expenditure from basal metabolic rate	total count
Marginal calories - Total marginal estimated energy expenditure	kilocalories
Resting heart rate	bpm

TABLE V
LIST OF THE MOST RELEVANT FEATURES DERIVED FROM THE SEQUENTIAL FORWARD SELECTION ALGORITHM

Feature name	Description
Mean_diff	Mean of differences of the signal, calculated for the last hour prior to sleep
Median_diff	Median of differences of the signal, calculated for the last hour prior to sleep
Centroid	Centroid along the time axis, calculated for the last couple of hours prior to sleep
Spectral_distance	Single spectral distance, calculated for the last 8 hours prior to sleep
Spectral_decrease	Amount of decreasing of the spectral amplitude, calculated for the last 11 hours prior to sleep
Wavelet_absolute_mean	CWT absolute mean value of wavelet scale, calculated for the last 10 hours prior to sleep
ECDF_Percentile	Percentile values of ECDF, calculated for the last 3 hours prior to sleep
MFCC	MEL cepstral coefficients, calculated for the last hour prior to sleep
FFT_mean_coefficients	Mean values of each spectrogram frequency, calculated for the last 1, 2, 5, 6, 8, 9 10 and 12 hours prior to sleep
TBR	Percentage time below range, [3, 3.9) mmol/L, calculated for the last 3 hours prior to sleep

task at hand and so, for all features to contribute equally to the result, same range is essential. On the contrary, this is not true for tree-based algorithms. Since a decision tree is only splitting a node based on a single feature, it is insensitive to the scale of other features. Such an example is the Random Forest algorithm which we also use in this work.

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

A standard 80-20 random split was used to separate the dataset into training set (80%) and holdout set (20%). To tune the hyperparameters of the two binary classifiers (SVM and Random Forest) and assess their effectiveness, a 10-fold stratified Cross Validation was implemented in the training set. During a k-fold Cross Validation procedure the training dataset is partitioned into k equal groups (folds) and each time training takes place in the k-1 folds and testing in the remaining one. Generally Cross Validation results in a less biased and less optimistic model performance. The primary metric used to evaluate the performance of the two algorithms was ROC-AUC, along with Sensitivity and Specificity (2), (3), where TP, TN, FP, FN are True Positives, True Negatives, False Positive and False Negatives respectively. A separate validation was also performed in the holdout set.

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{2}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{3}$$

III. RESULTS

The performance of the two classifiers after a 10-fold cross validation in the training set and in the holdout is shown in Table VI. Both algorithms seem to generalize quite well in new unseen data. However, we see that SVM outperforms Random Forest classifier, especially due to the poor performance of the latter in terms of the Specificity metric. On the other hand, SVM is quite balanced across all metrics. In Figure 1, we can also see the ROC curve of the SVM algorithm for the Training and Holdout set accordingly.

The best predictors proved to be glucose related patterns (as shown in Table V) and thus these features were included in the model. Demographics such as age, sex and BMI did not add significant value in predicting NH. This may be due to the restriction of range, because of the relatively small number of participants in the dataset. By focusing on a small group of participants (37 in this clinical trial), we have necessarily limited our ability to investigate all possible variables. This finding was expected, and it does not necessarily contradict with other studies which have declared demographic features among the most important ones for the task of NH prediction.

Features related to administered insulin and consumed carbohydrates did not contribute equally to glucose features either. The fact that participants were asked to manually log insulin units and meal information has led to some inconsistencies in the data. Specifically, some unrealistic values due to human error as well as outdated or missing information because users might have forgotten to report accordingly could have led to noise in the dataset which does not allow the algorithm to find useful patterns.

Finally, even though many studies have pointed out the importance of physical activity as one of the main risk factors in the occurrence of NH events [16] (especially evening exercise), in this work we did not find such a strong relationship. One explanation might be that the effects of physical activity as well as insulin and meal information are already included in the glucose-related features used by the model, making these specific features less significant. However, this is just a hypothesis, and no further evidence is available at the moment.

In Table VII, we show the performance of the SVM classifier in a separate dataset of a completely unseen population. This population consists of 560 nights, where 20 adult participants were wearing a CGM Dexcom G5. The ROC-AUC seems to be reasonably high with just a small deviation from the holdout set, which implies the robustness and ability of the algorithm to generalize in a new populations data. Furthermore, we tested the algorithm in the same population, this time while using FreeStyle Libre for an additional 551 nights. The performance of the algorithm in terms of the ROC-AUC remains high, which highlights its ability to be used with devices from different manufacturers, providing significant value in the discrimination between the nights with or without hypoglycaemia.

In Figure 2 we can see the ROC curve along with the Recall-Precision curve of the proposed SVM classifier, evaluated in the dataset acquired from the flash sensors. It is worth noting that the classifier was trained to optimize the ROC-AUC score and hence its Precision does not remain always at a high level. Specifically, at best we can adjust the predictive threshold so that the model predicts 45% of the nights with NH, with more than 70% Precision, while misclassifying only 10% of the nights with absence of any NH event.

TABLE VI

AUC-ROC SCORE, SENSITIVITY AND SPECIFICITY OF THE TWO MACHINE LEARNING CLASSIFIERS, EVALUATED DURING A 10-FOLD CROSS VALIDATION IN THE TRAINING AND IN THE HOLDOUT SET.

	SVM (kernel: poly, gamma: scale, C:0.1)		Random Forest (n_estimators: 50)	
	Training Set (10-fold CV)	Holdout Set	Training Set (10-fold CV)	Holdout Set
ROC-AUC	79.36% (95% CI, 76.86%-81.86%)	78.51%	75.80% (95% CI, 74.00%-77.60%)	76.96%
Sensitivity	73.66% (95% CI, 70.36%-76.96%)	72.13%	42.69% (95% CI, 39.19%-46.19%)	43.88%
Specificity	72.31% (95% CI, 70.51%-74.11%)	71.71%	87.50% (95% CI, 86.60%-88.40%)	89.31%

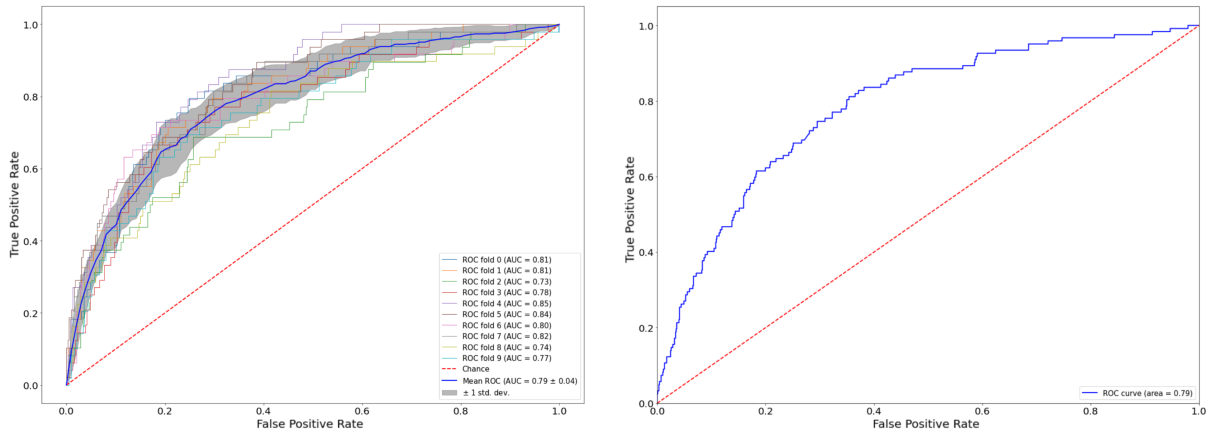

Fig. 1. Receiver operating characteristics curve from the evaluation of the SVM binary classifier, using a Cross Validation in the training set (left) and in the holdout set (right).

TABLE VII

EVALUATION OF THE SVM ALGORITHM IN A NEW UNSEEN DATASET.

	CGM (Dexcom G5)	Flash glucose monitoring (Abbott FreeStyle Libre)
ROC-AUC	77.06%	77.74%
Sensitivity	73.76%	82.61%
Specificity	65.63%	56.69%

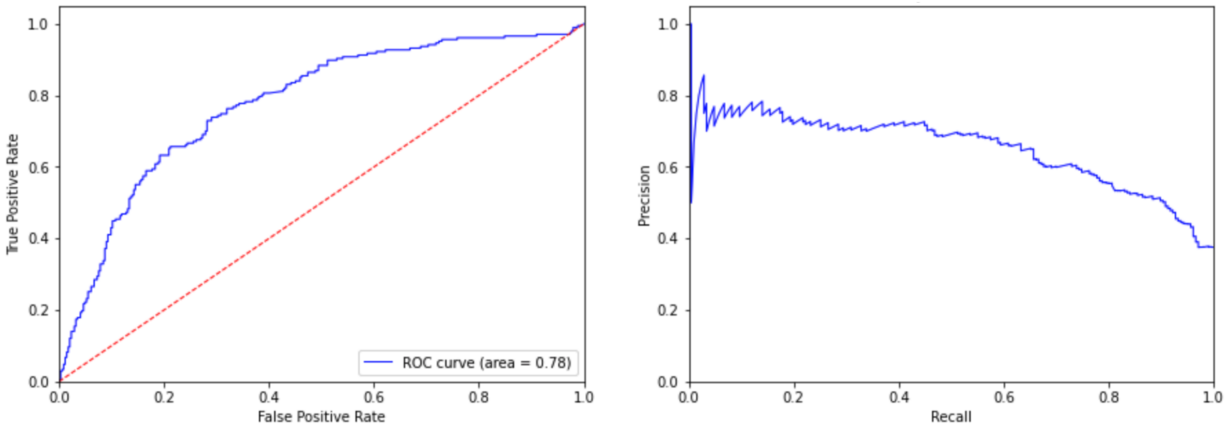

Fig. 2. ROC curve (left) and Recall-Precision curve (right) of the proposed SVM binary classifier evaluated in a dataset of 20 adult participants wearing flash glucose monitoring sensor during a period of 8 weeks.

IV. Discussion

In this work we investigate an open-source feature engineering library [12] for time series, to capture as much as possible of discriminative signal characteristics of continuous glucose data. We show that this is a feasible approach for predicting NH before bedtime and can also be used for the prediction of other adverse events (such as hyperglycaemia). In our training dataset, SVM algorithm seems to outperform Random Forests with an AUC score of 79%, which is among the state-of-the-art performances in the related literature. SVM is intrinsically a binary classification problem which maximizes the margin between the two classes by calculating the distance between different points. So, for n points in a dataset it calculates n^2 dot products and therefore is hardly scalable for very large datasets. However, for smaller ones it has proven to perform better than Random Forests in many other applications as well.

We have seen that the algorithm utilizes glucose-related features to make its predictions, without the need from other data sources (administered insulin, meal intake or physical activity). This is an advantage of the proposed model especially if we think about providing bedtime decision support to people with T1D, to minimize the risk of NH. Specifically, if this model is implemented in an existing CGM/flash system or as a standalone decision support system for diabetes management [17]–[19], it can provide help to people with fear of NH, by informing them of a potential risk and advising them to take appropriate action (reduction of basal insulin units and/or a meal of complex carbohydrates before bedtime). This can be achieved without the need for extra data sources (I.e., meal, insulin, physical activity) which might be unavailable, incomplete and not always accurate.

The generalization power of the SVM classifier was further tested in a completely new population. We show that its ability to discriminate remains reasonably high, however we observe a drop in Specificity. One reason for that may be the consumption of rescue carbohydrates before bedtime, used by the participants as an intervention to prevent NH. Even though for the training of our algorithm we excluded such cases, meal information was not available in the second test set. As such, glucose levels might have been impacted from snacks before night-time sleep, leading to an increased number of False Positives. Another reason may be that the testing dataset comes from a high-risk population with frequent previous severe hypoglycaemic events or impaired hypoglycaemia awareness. Specifically, the prevalence of NH events is 3-fold higher compared with the training set. Even though in theory Sensitivity and Specificity metrics should not be affected by this (as opposed to Positive Predictive Value and Negative Predictive Value), a dip especially in Specificity has been associated with an increase in the incidence of an event or the prevalence of a disease in many other studies [20].

The testing set consists of 20 high-risk adult participants while wearing a CGM Dexcom G5 for 8 weeks and an FreeStyle Libre 1 for an additional 8 weeks. In both cases the model performed equally well in terms of its discriminative capabilities, as presented in Table 2. To the best of our knowledge, this is the first study of its kind providing evidence that a NH prediction algorithm is tolerant of differing sampling frequencies and accuracy characteristics, it can be device agnostic and perform well on data acquired from sensor devices produced by different manufacturers. It has been shown that the role of flash glucose monitoring in the self-management of T1D is less clear, especially for people with impaired awareness of hypoglycaemia [10], [21], [22]. So, algorithms like the one proposed in this work reveal an opportunity to enhance the functionality of flash monitoring systems along with their CGM counterparts.

Due to the high variability of blood glucose response in people with T1D, it is possible that one algorithm cannot cover the whole spectrum of population characteristics. As such, transfer learning might help to learn individual blood glucose patterns before making future predictions. So, the current algorithm can be used during an initialization period (for example 2 weeks), where conservative treatment is offered as intervention based on the clinical guidelines. After this period the model gets retrained in each individual's data to become more personalized, learn individual blood glucose patterns and as a result be able to make more informed predictions and provide better recommendations to individuals. However, nocturnal hypoglycaemia prevention was not the aim of this work and thus further research needs to be done in this direction.

V. Conclusions

In this work we propose an extensive feature engineering machine learning framework for feature extraction from glucose time series. We train binary classifiers to predict before bedtime the occurrence of nocturnal hypoglycaemia events. The proposed model is an SVM algorithm using glucose-related features and achieving a ROC-AUC score of 79%. We prove our models generalizability by testing it in a completely unseen population of high-risk adults with T1D. We observe a dip in Specificity, however the problem at hand is mainly recall-oriented and so the focus is on identifying NH nights en masse. Implications of precision errors made by the model on clinical outcomes have been studied in the literature, and they seem to be outperformed by the benefits of an early diagnosis. We also show that our model is device agnostic and hence an integration with a decision support system has the potential to reach a wider range of users, informing them about their risk of NH before it occurs.

References

[1] Danne T, Nimri R, Battelino T, Bergenstal RM, Close KL, DeVries JH, Garg S, Heinemann L, Hirsch I, Amiel SA, Beck R. International consensus on use of continuous glucose monitoring. Diabetes care. 2017 Dec 1;40(12):1631-40.

[2] Graveling AJ, Frier BM. The risks of nocturnal hypoglycaemia in insulin-treated diabetes. Diabetes research and clinical practice. 2017 Nov 1;133:30-9.

[3] Seaquist ER, Anderson J, Childs B, Cryer P, Dagogo-Jack S, Fish L, Heller SR, Rodriguez H, Rosenzweig J, Vigersky R. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. The Journal of Clinical Endocrinology & Metabolism. 2013 May 1;98(5):1845-59.

[4] Vu L, Kefayati S, Idé T, Pavuluri V, Jackson G, Latts L, Zhong Y, Agrawal P, Chang YC. Predicting Nocturnal Hypoglycemia from Continuous Glucose Monitoring Data with Extended Prediction Horizon. InAMIA Annual Symposium Proceedings 2019 (Vol. 2019, p. 874). American Medical Informatics Association.

[5] Jensen MH, Dethlefsen C, Vestergaard P, Hejlesen O. Prediction of nocturnal hypoglycemia from continuous glucose monitoring data in people with type 1 diabetes: a proof-of-concept study. Journal of diabetes science and technology. 2020 Mar;14(2):250-6.

[6] Bertachi A, Viñals C, Biagi L, Contreras I, Veh J, Conget I, Gimnez M. Prediction of Nocturnal Hypoglycemia in Adults with Type 1 Diabetes under Multiple Daily Injections Using Continuous Glucose Monitoring and Physical Activity Monitor. Sensors. 2020 Jan;20(6):1705.

[7] Güemes A, Cappon G, Hernandez B, Reddy M, Oliver N, Georgiou P, Herrero P. Predicting quality of overnight glycaemic control in type 1 diabetes using binary classifiers. IEEE journal of biomedical and health informatics. 2019 Sep 13;24(5):1439-46.

[8] Marling C, Bunescu R. The OhioT1DM dataset for blood glucose level prediction: Update 2020. InCEUR workshop proceedings 2020 Sep (Vol. 2675, p. 71). NIH Public Access.

[9] Mosquera-Lopez C, Dodier R, Tyler NS, Wilson LM, El Youssef J, Castle JR, Jacobs PG. Predicting and preventing nocturnal hypoglycemia in type 1 diabetes using big data analytics and decision theoretic analysis. Diabetes technology & therapeutics. 2020 Nov 1;22(11):801- 11.

[10] Reddy M, Jugnee N, El Laboudi A, Spanudakis E, Anantharaja S, Oliver N. A randomized controlled pilot study of continuous glucose monitoring and flash glucose monitoring in people with type 1 diabetes and impaired awareness of hypoglycaemia. Diabetic Medicine. 2018 Apr;35(4):483-90.

[11] Reddy M, Jugnee N, Anantharaja S, Oliver N. Switching from flash glucose monitoring to continuous glucose monitoring on hypoglycemia in adults with type 1 diabetes at high hypoglycemia risk: the extension phase of the I HART CGM study. Diabetes technology & therapeutics. 2018 Nov 1;20(11):751-7.

[12] Barandas M, Folgado D, Fernandes L, Santos S, Abreu M, Bota P, Liu H, Schultz T, Gamboa H. TSFEL: Time series feature extraction library. SoftwareX. 2020 Jan 1;11:100456.

[13] McCall AL, Cox DJ, Crean J, Gloster M, Kovatchev BP. A novel analytical method for assessing glucose variability: using CGMS in type 1 diabetes mellitus. Diabetes technology & therapeutics. 2006 Dec 1;8(6):644-53.

[14] Wilinska ME, Chassin LJ, Schaller HC, Schaupp L, Pieber TR, Hovorka R. Insulin kinetics in type-1 diabetes: continuous and bolus delivery of rapid acting insulin. IEEE Transactions on Biomedical Engineering. 2004 Dec 20;52(1):3-12.

[15] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over- sampling technique. Journal of artificial intelligence research. 2002 Jun 1;16:321-57.

[16] Riddell MC, Gallen IW, Smart CE, Taplin CE, Adolfsson P, Lumb AN, Kowalski A, Rabasa-Lhoret R, McCrimmon RJ, Hume C, Annan F. Exercise management in type 1 diabetes: a consensus statement. The lancet Diabetes & endocrinology. 2017 May 1;5(5):377-90.

[17] Tyler NS, Mosquera-Lopez CM, Wilson LM, Dodier RH, Branigan DL, Gabo VB, Guillot FH, Hilts WW, El Youssef J, Castle JR, Jacobs PG. An artificial intelligence decision support system for the management of type 1 diabetes. Nature metabolism. 2020 Jul;2(7):612-9.

[18] Mougiakakou SG, Bartsocas CS, Bozas E, Chaniotakis N, Iliopoulou D, Kouris I, Pavlopoulos S, Prountzou A, Skevofilakas M, Tsoukalis A, Varotsis K. SMARTDIAB: a communication and information technology approach for the intelligent monitoring, management and follow-up of type 1 diabetes patients. IEEE Transactions on Information Technology in Biomedicine. 2010 Jan 29;14(3):622-33.

[19] Liu C, Avari P, Leal Y, Wos M, Sivasithamparam K, Georgiou P, Reddy M, Fernndez-Real JM, Martin C, Fernndez-Balsells M, Oliver N. A modular safety system for an insulin dose recommender: a feasibility study. Journal of diabetes science and technology. 2020 Jan;14(1):87-96.

[20] Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. Cmaj. 2013 Aug 6;185(11):E537-44.

[21] Hásková A, Radovnická L, Petruželková L, Parkin CG, Grunberger G, Horová E, Navrátilová V, Kádě O, Matoulek M, Prázný M, Šoupal J. Real-time CGM is superior to flash glucose monitoring for glucose control in type 1 diabetes: the CORRIDA randomized controlled trial. Diabetes Care. 2020 Nov 1;43(11):2744-50.

[22] Visser MM, Charleer S, Fieuws S, De Block C, Hilbrands R, Van Huffel L, Maes T, Vanhaverbeke G, Dirinck E, Myngheer N, Vercammen C. Comparing real-time and intermittently scanned continuous glucose monitoring in adults with type 1 diabetes (ALERTT1): a 6-month, prospective, multicentre, randomised controlled trial. The Lancet. 2021 Jun 12;397(10291):2275-83.