Temporal Specialization in LSTM Models for Enhanced Prediction of Adverse Glycaemic Events in Type 1 Diabetes

Abdelaziz Mansour[1*] , Kamal Amroun[1] , Zineb Habbas[2]

[1] LIMED Laboratory, Faculty of Exact Sciences, University of Bejaia, Bejaia 06000, Algeria
[2] LORIA Laboratory, Lorraine University, Vandœuvre-lès-Nancy 54506, France

Corresponding Author Email: abdelaziz.mansour@univ-bejaia.dz

ABSTRACT

This work focuses on the prediction of adverse events (hypoglycaemia and hyperglycaemia) in Type 1 Diabetes (T1D) patients, with a prediction horizon (PH) of 30 minutes. Utilizing an 8-week dataset with blood glucose (BG) measurements every 5 minutes from 12 T1D patients, the approach adopted consists of developing specific models for each time period: diurnal, nocturnal and global (including both periods), recognizing that factors influencing BG levels can vary considerably between the daytime and night-time periods. The algorithm used for this study is the Long Short-Term Memory (LSTM), known for its ability to efficiently process complex temporal sequences. The models are multi output, they provide predictions over PHs of 5, 10, 15, 20, 25 and 30 minutes simultaneously which are then combined to have the final prediction for a PH of 30 min. The results from the global models, based on full dataset, reveal disparate performances between diurnal and nocturne periods. This disparity highlights the importance of exploring a finer approach by creating specific models for each temporal period. The best F1-scores obtained for the daytime and night-time hypoglycaemia models are 0.609 and 0.675, respectively, while they are 0.838 and 0.866 for the daytime and night-time hyperglycaemia models, respectively.

1. INTRODUCTION

For individuals with Type 1 Diabetes (T1D), managing blood glucose (BG) levels is a daily challenge, as it requires careful monitoring and insulin administration to avoid both hyperglycaemia (high BG levels) and hypoglycaemia (low BG levels). Traditionally, this was achieved through self-monitoring of blood glucose (SMBG) using finger-stick measurements, performed a few times a day. However, with the advent of Continuous Glucose Monitoring (CGM) devices, patients can now track their BG levels almost continuously. CGM systems provide real-time BG readings, generating continuous streams of data that capture fluctuations in BG levels throughout the day. This data-rich environment enables machine learning (ML) and deep learning (DL) models to analyse patterns and trends in BG dynamics, thereby making it possible to forecast future BG levels. By incorporating these models into edge devices such as insulin pumps or CGM, patients can gain vital insights, allowing them to better manage their BG levels and take proactive measures to prevent complications.

Considerable research has been dedicated to developing and refining BG prediction models, exploring a wide range of approaches, including classical statistical methods, autoregressive models, and cutting-edge DL architectures. These models can account for multiple factors affecting BG levels, such as meal intake, physical activity, stress, and insulin dosages.

Despite the progress, challenges remain in fine-tuning these models to handle the inherent variability in BG responses across different individuals and circumstances. A major challenge in developing more accurate prediction models is addressing the considerable variation in BG behaviour between daytime and night-time periods. Indeed, prediction models found in the literature that are used to make 24-hour cycle predictions can often lead to misleading results due to the failure to distinguish between daytime and night-time performance. It is essential to acknowledge that daytime and night-time conditions pose distinct challenges for prediction models, which can greatly affect their accuracy and reliability.

BG behaviour in patients with T1D shows significant variations throughout the day, influenced by daytime and night-time periods. During the day, physical activity levels, meals and various environmental factors can have a direct influence on BG levels. Meals, in particular, induce an increase in BG as carbohydrates are metabolized, thus requiring precise adjustment of the insulin dose to maintain stable BG levels. Physical exercise, on the other hand, can cause a decrease in BG, requiring careful management of insulin intake to avoid hypoglycaemia. At night, the challenges of BG control evolve. Periods of sleep may be associated with unpredictable fluctuations in BG levels, influenced by factors such as the release of certain hormones, decreased physical activity, and the length of night-time fasting.

The separate evaluation of daytime and night-time

performance should allow a more precise analysis of the strengths and weaknesses of the models. Models that excel in daytime conditions may not perform as well at night, and vice versa. Therefore, an overall performance assessment may mask specific shortcomings and hinder a thorough understanding of the model's true ability to adapt to diverse situations.

Our methodology approaches adverse event prediction as a binary classification problem. The aim is to detect adverse events using "Leave One Patient Out Cross Validation". We applied a univariate approach that focuses exclusively on BG measurement data collected by CGMs. The literature suggests that including additional features, such as insulin boluses, provides only minor enhancements in performance.

As part of this work, we opted for the use of Long Short-Term Memory (LSTM). LSTMs are particularly suited to modelling complex temporal sequences, making them a good choice for our data analysis.

The particularity of our approach lies in the use of a multi-output LSTM models. In other words, our models can simultaneously generate six predictions corresponding to six different Prediction Horizons (PHs): 5, 10, 15, 20, 25 and 30 minutes. It should be emphasized that the first layers of the LSTM network are common to all PHs, thus facilitating the learning of general characteristics inherent in the data. In contrast, the final layers of the network are specific to each PH, allowing the model to capture finer, specialized contextual information.

Furthermore, one of the crucial steps in our approach was the division of the overall dataset into three distinct subsets, each representing a specific temporal aspect. These sub-datasets are as follows:

- **Diurnal dataset:** we isolated diurnal data from the full dataset.
- **Nocturne dataset:** we extracted all data associated with the night-time period from the full dataset.
- **Complete dataset:** we kept the overall dataset, which encompasses both daytime and night-time data.

Indeed, the contributions presented in this work are:

- Analysis of the performance of adverse event prediction models depending on the period: daytime or night-time.
- Integration of an approach based on a multi-output model that makes predictions on six PHs simultaneously.

The rest of this article is structured as follows: The first section focuses on diabetes, exploring its essential aspects. The second section reviews previous research that has addressed the issue of BG prediction in patients with TID using artificial intelligence-based approaches. The third section details our methodology, highlighting the preparation of the dataset and the development of the models used in our study. Sections 6 and 7 are dedicated to the presentation of the results obtained, followed by an in-depth discussion. Finally, the last section offers a synthetic conclusion, summarizing the key points covered in the article.

2. DIABETES

Diabetes, as a chronic disease, represents a major public health challenge in the world [1]. There are several types of diabetes, among which the most common are T1D, Type 2 Diabetes (T2D) and gestational diabetes. Each of these types has distinct characteristics, but they all share one thing: a failure in the BG regulation system. This irregularity results from a failure in the production of an essential hormone secreted by the beta cells of the pancreas, called *insulin*. The latter allows the body's cells to absorb glucose present in the blood. In the absence of insulin, glucose builds up in the blood, leading to serious complications.

T1D is usually diagnosed in children and young adults. In this case, the immune system attacks the beta cells of the pancreas. T2D, which is more prevalent in adults, is distinguished by insulin resistance or insufficient production of this hormone. Thus, cells become less responsive to insulin, making it difficult for cells to absorb glucose, contributing to an increase in BG levels. Gestational diabetes, on the other hand, occurs during pregnancy and can increase the risk of complications for both mother and child. Although this type of diabetes is often temporary, it requires close monitoring to avoid complications during pregnancy and reduce the later risk of developing T2D.

Diabetes management is based on approaches specific to each type of diabetes. Treatment of T2D and gestational diabetes may involve oral medications, insulin injections or other medications, lifestyle changes including eating a balanced diet and regular physical activity. People with T1D rely on insulin, which they give by injection. Two types of insulin are generally used: basal insulin and bolus insulin. Basal insulin aims to maintain a constant level of insulin in the body, working in the background to cover basic needs between meals and during the night. Alternatively, bolus insulin is administered before meals to control the increase in BG caused by food ingestion. Patients often use a set of devices, including insulin pumps and CGM systems. Insulin pumps are portable electronic devices that provide precise and flexible insulin delivery. They eliminate the need for frequent injections. On the other hand, CGM systems are small sensors implanted under the skin that continuously measure the concentration of glucose in interstitial fluid, providing patients with real-time information about their BG levels.

3. RELATED WORKS

The problem of predicting BG levels in T1D patients has been widely studied in the literature. Particularly in recent years, with the appearance of new technologies such as: CGMs and insulin pumps. A lot of information having an impact on BG variability was collected, via dedicated devices, such as: BG values measured with regular frequency, basal insulin injection rate, boluses of insulin, heart rate measurement, number of steps taken, body temperature measurements, etc. This collected information, assimilated to time series, was used to construct datasets: privates [2-4] or public [5, 6]. These data were then used by data-driven techniques to predict future BG variability, generally over a short time horizon ranging from 15 minutes to 60 minutes [2, 7-9]. Indeed, fewer studies focus on predictions longer than 60 minutes [10-12].

Artificial intelligence techniques based on ML and DL have been widely used. ML is employed through various algorithms, including Linear Regression algorithms [3, 13], Random Forest (RF) [2, 11], Support Vector Machine (SVM) [12, 14-16]. As for DL, the algorithms used are: Recurrent neural networks (RNN) [17], Convolutional Neural Networks (CNN) [18], LSTM [7, 9, 19], Gated Recurrent Unit (GRU) [20].

The literature offers various approaches, characterized by several distinctions mainly as outlined below:

- Univariate models or multivariate models

- Regression or classification
- Binary classification or classification with more than 2 classes
- Sample-based or event-based prediction
- Precision medicine, leave one patient out Cross Validation or hold out validation

The univariate approach refers to considering only one characteristic as input at a time (example: values of past CGMs) [3, 7, 9, 17], while multivariate approaches refer to simultaneously considering several characteristics as inputs (example: previous information on BG, insulin boluses, CHO and physical activity, etc.) [11, 19, 21]. Regression is used to predict the continuous and accurate BG value within a given PH (usually 30 minutes) [9, 17, 22, 23]. The classification, on the other hand, is used to predict which category the BG will belong to in a given PH, for example, whether its BG will be normal, high or low, without specifying the exact continuous value [2-4, 11]. Binary classification encompasses the categorization into two distinct classes (presence or absence of an adverse event such as hypoglycaemia) [2, 4], and multi-class classification encompasses categorization into three or more classes (Example: Severe Hypoglycaemia, Hypoglycaemia, Normal, Hyperglycaemia, Severe Hyperglycaemia) [7]. For regression, evaluation metrics such as mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) are used, while in classification, the evaluation is based on precision, recall, and F1-score.

Two distinct approaches in the context of predicting glycaemic events. The sample-based approach consists of making predictions for each measurement time, autonomously classifying each data sample according to the PH. On the other hand [2, 7], the event-based approach groups together several consecutive measurement times indicating a similar glycaemic state in a single event. Predictions focus on the occurrence of a glycaemic event within the next few minutes, and a predicted event is considered a True Positive (TP) if it actually occurs within that time frame [3, 24].

Precision medicine [7, 20, 21, 25] recognizes that each patient has unique medical characteristics and responses. This involves the development of a model tailored specifically to each patient, while Leave One Patient Out Cross Validation [15, 24] consists of training the model on the data of all patients except one, then carrying out the evaluation on this excluded patient. This procedure is repeated until each patient has been isolated at least once for use as a test set. This helps simulate how the model generalizes to data from patients it has not encountered during training.

This study aims essentially to evaluate the effectiveness of models in predicting adverse events during distinct temporal periods, specifically diurnal or nocturnal, and compare their performance against models using a comprehensive dataset covering both periods. To our knowledge, there is limited existing research addressing this aspect, with only a few works identified [26, 27].

Table 1. Overview of studies of glycaemic event prediction models in the literature

Work	Dataset # Patients	Validation	Model	Period	Approach	Features	PH	/	Results Hypo	Results Hyper
[4]	Private, 112 T1D patients	Precision medicine	LSTM	All day	Event-based	Univariate CGM data	30 min	S	92.50	/
								R	92.59	/
[24]	Ohio1dm, 12 T1D patients	Leave one patient cross validation	LSTM, CNN, CART Decision Tree	All day	Event-based	Univariate CGM data	30 min, α = 1	P	97.5	79.7
								R	81,3	51.4
								F1	88.7	62.5
[14]	Private, 10 T1D patients	Precision medicine	SVM	Nocturne	Sample-based	Multivariate	6 hours	S	82.15%	/
								R	78.75%	/
[12]	Private, 37 T1D patients for train and 20 for test	Hold-out validation	SVM	Nocturne	Sample-based	Multivariate	6 hours	S	65.63%	/
								R	73.76%	/
[28]	Private, 21 T1D and 10 T2D	5-fold cross-validation	CART	All day	Event-based	Univariate CGM data	15 min	S	79.9%	/
								R	80.05%	/
[15]	Private, 10 T1D	leave-one-subject-out cross-validation	SVM	All day	Event-based	Multivariate	Real time	R	100%	/
[2]	Private, 104 patients (52 T1D and 52 T2D)	5-fold cross-validation	RF	Postprandial	Sample-based	Multivariate	30 min	S	91.3%	/
								R	89.6%	/
								F1	0.543	/
[16]	Private, 10 T1D	Precision medicine	SVC	Postprandial	Sample-based	Multivariate	240 min	S	79%	/
								R	71%	/
[29]	Private, 10 T1D	Holdout validation	ANN	Nocturne	Event-based	Multivariate ECG data	Real time	S	/	65.38%
								R	/	70.59%
[30]	Public, 463 T1D	Fivefold cross-validation	Linear Discriminant Analysis (LDA)	Nocturne	Sample-based	Multivariate	All night	S	70%	/
								R	75%	/

Note: S: Specificity, R: Sensitivity. F1: F1-Score. When several models are used in a study, the model that yielded the best results is mentioned in this table

The study [31] present approaches for predicting adverse events over a 24-hour period. On the other hand, the scientific literature includes works that has focused on specific aspects of the prediction of adverse events within well-defined temporal periods. For instance, some have explored the prediction of nocturnal hypoglycaemia / hyperglycaemia, as evidenced in works [8, 11, 12, 14, 28]. Additionally, others have focused on predicting adverse events in specific intervals, such as the postprandial period [2, 16]. The study [26] evaluates the effectiveness of a Support Vector Regression (SVR) model to predict night-time and daytime hypoglycaemic events separately over 30 and 60 minute periods. Using data from 15 T1D patients containing information on recent glycaemic profiles, meals, insulin and physical activities. The results obtained highlight the need to approach hypoglycaemia prediction differently for night-time and daytime periods. The approach proposed in the studies [11, 12] relies on the analysis of data commonly collected during the day, such as CGM data, food intake, and insulin boluses to predict the quality of night-time glycaemic control in people with T1D. The works [2, 16] aims to evaluate the feasibility of a ML algorithm to predict hypoglycaemia over the entire postprandial period (240 minutes) in people with T1D. In the study [16], the Support Vector Classifier (SVC) was used where two levels of hypoglycaemia were defined based on their severity (Level 1: Glucose $\leq$ 70 mg/dL and Glucose $\geq$ 54, Level 2: Glucose < 54 mg/dL). The evaluation was carried out on real data from 10 patients, with personalized models for each individual. The results suggest that it is possible to anticipate postprandial hypoglycaemic events with a reasonable rate of False Positives (FPs), making it possible to avoid more than two thirds of these events. The study [2] used a dataset including 104 type 1 and type 2 diabetic patients. They evaluated 4 different models to predict short-term postprandial hypoglycaemia over a 30-minute period. The results indicated that the RF-based model outperformed the others in terms of predictive performance. Indeed, the

attention given to the prediction of hypoglycaemia is notably higher than that dedicated to hyperglycaemia. This could be explained by the increased risks associated with hypoglycaemia, including neurological complications and potentially serious short-term consequences. It is also notable that many articles treat these two events jointly, while fewer research efforts focus specifically on hyperglycaemia. For example, the study [29] developed a classification unit based on a multilayer forward-propagating neural network, using electrocardiographic (ECG) parameters for the prediction of hyperglycaemia. The study [24] employed an event-based approach where they introduced a layered meta learning algorithm. Their approach utilizes LSTM or CNN as base learners, and a decision tree as meta learners. They employed a parameter α to assess the level of advancement in the predictions, examined over a range from 1 to 6 (corresponding to a period of 5 to 30 minutes). The approach is based exclusively on univariate data, specifically CGM data collected throughout the entire day. In addition to detecting hypoglycaemia and hyperglycaemia, the study also aims to detect normoglycaemia events with a PH of 30 minutes. The study includes an evaluation of the practical feasibility of the approach in real-world scenarios by implementing it on an embedded device.

Table 1 provides an overview of the state of the art in predicting glycaemic events.

4. MATERIAL AND METHOD

In this section, we present the methodology adopted to carry out our study. This includes the description of the dataset used, the different pre-processing steps applied, defining the input variables and the outcome labels, as well as the modelling and the parameter search procedure. Figure 1 presents an overview of our approach, outlining the fundamental aspects involved in the development and evaluation of the predictive model.
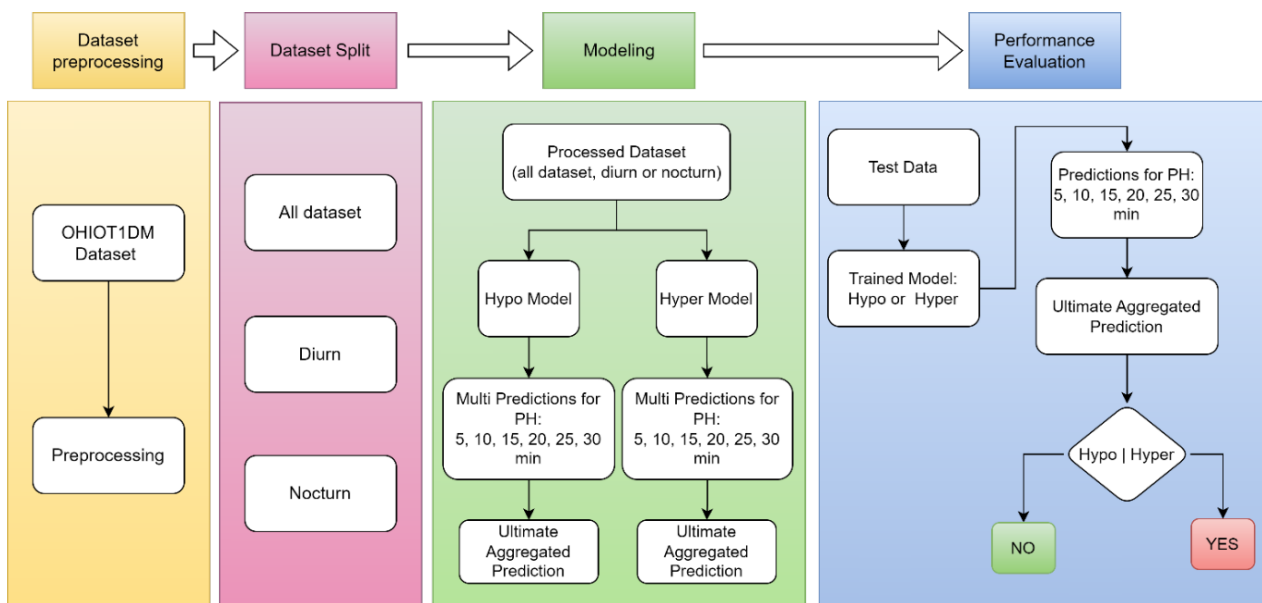

Figure 1. Schematic overview for model prediction development and assessment

4.1 Description of the OHIOTDM dataset

To evaluate the effectiveness of our approach, this study relies on data from OhioT1DM [5]. This dataset was

specifically designed to be used for the development of data-driven models for predicting BG levels in T1D patients. The data was collected from 12 patients over a period of 8 weeks (data from 6 patients were made public in two in 2018,

followed by 6 others in 2020). Each patient's dataset is divided into two subsets, reserving the last week of data for testing purposes. The features of the dataset show diversity, with variables such as BG measurements, insulin bolus, meal information, sleep information, exercise intensity, time of self-reported stress, illness, and hypoglycaemic events. A second category includes CGM, basal insulin level, galvanic skin response, heart rate, air and skin temperature, and step counter data.

4.2 Pre-processing

The development of prediction models is initiated with rigorous pre-processing steps performed on the dataset. This notably includes handling missing values and normalizing variables. The pre-processing steps are detailed as follows:

Converting XML files to CSV format: we converted the files provided in XML format to CSV format for easier further processing.

CGM data extraction: we collected all the data relating to CGM measurements, serving as inputs for our prediction model.

Filling CGM missing data: we initially replaced missing data points with values calculated by interpolating one step forward and one step backward in time. This method helps to estimate missing values based on the available data before and after the gaps, providing a more accurate representation of the BG levels during those time periods. However, any remaining missing CGM values were subsequently filled with zeros.

To ensure data quality, we implemented a rule regarding missing data rates. If a specific day or night-time period in the dataset exhibited a missing data rate greater than 33%, we opted to remove all data corresponding to that day or night. This strategy was employed to prevent potential bias and inaccuracies that could arise from significant gaps in the data. While this decision may affect the overall size of the dataset, it enhances our confidence in the integrity of the remaining data.

Merging training and testing data: we combined the training and testing datasets for each patient as we used the Leave one patient out cross validation approach by leaving one patient out in each iteration.

Splitting into diurnal and nocturnal sub-datasets: to define the specific time periods chosen for the diurnal and nocturnal models, we focused on leveraging the "sleep" and "basis_sleep" features available in the dataset. These features provided detailed information about when patients were asleep, allowing us to split the dataset into two distinct time periods: diurnal and nocturnal.

- **"Sleep" feature:** this data is reported directly by the patient, indicating the precise times when the patient starts and ends their sleep. It gives insight into the actual sleep habits of each individual, which may vary significantly depending on personal routines and lifestyle factors.
- **"Basis_sleep" feature:** this feature comes from the dedicated sensor tracking the patient's physiological state, automatically detecting when the patient is asleep. It provides a reliable, sensor-based indication of the patient's sleep schedule, which complements the self-reported "sleep" feature.

To identify the most accurate sleep window, we take the intersection of the two periods by combining the last sleep time with the earliest wake-up time from both features.

In cases where both the "sleep" and "basis_sleep" features

were unavailable throughout the night, we standardized the sleep period by defining it from 10:00 p.m. to 6:00 a.m. This choice is based on the typical sleep habits observed among patients in the dataset. Many patients in the dataset follow a routine that aligns with these hours, making it a reasonable approximation for the nocturnal period when specific sleep data is missing. Accordingly, the **diurnal period** was defined from the patient's awakening until the onset of sleep, representing the active, wakeful hours of the day.

Normalization of CGM values: to enhance the model's ability to identify patterns and relationships in the data, we applied normalization to the CGM values. Specifically, we used the min-max scaling technique, which transforms features to a fixed range between 0 and 1. This is essential because LSTM models typically perform better and train faster when input features are on a uniform scale. Large differences in feature magnitudes can lead to instability during training. Furthermore, normalization accelerates the convergence of gradient descent-based optimizers, such as the Adam optimizer (Adaptive Moment Estimation) used in this study.

4.3 Establishing input variables and outcome labels

Each data entry in the dataset for a given time point includes the current CGM value at that specific time point x_t, as well as a sequence of previous CGM values within a time window of N timestamps $[x_{t-1}, x_{t-2}, ..., x_{t-(N-1)}]$.

Our method consists of evaluating the presence of an adverse event during the PH, rather than looking for the event precisely at the time $t + PH$. Thus, we defined a binary variable y_{t+PH} indicating the presence (1) or absence (0) of the adverse event in the time interval $[t + 1, t + PH]$.

If the adverse event is defined as hypoglycaemia, the threshold is set at 70 mg/dL. The risk label is defined as follows:

$$y_{t+PH} = \begin{cases} 1 & if\ \exists \hat{t} \in [t+1, t+PH]: x_{\hat{t}} < 70 \\ 0 & else \end{cases} \tag{1}$$

And similarly, if the adverse event is defined as hyperglycaemia, the threshold is set at 180 mg/dL and the risk label is defined as follows:

$$y_{t+PH} = \begin{cases} 1 & if\ \exists \hat{t} \in [t+1, t+PH]: x_{\hat{t}} > 180 \\ 0 & else \end{cases} \tag{2}$$

The selection of thresholds at 70 mg/dL for hypoglycaemia and 180 mg/dL for hyperglycaemia aligns with the recommendations of the American Diabetes Association (ADA) [32].

The optimal value of N was identified as 6 timestamps (30 minutes) after following experimental tests which aimed to evaluate the performance of the model under different time window configurations.

In this study, we set the PH to 30 minutes. Demonstrations have shown that this duration is effective in initiating corrective measures to deal with the event [33]. Furthermore, it provides the best compromise between minimizing error in prediction outcomes (the longer the PH, the greater the error) and preserving forecast accuracy [3].

To reduce FPs due to the risk of triggering alerts or actions based on temporary fluctuations or unusual single measurements that may not be clinically meaningful; An event or prediction is considered valid if it lasts for at least 10

minutes, corresponding to a minimum of three consecutive timestamps.

4.4 LSTM

In this work, we adopted an approach based on DL, more precisely RNNs with an LSTM architecture in order to predict adverse events, specifically hypoglycaemia and hyperglycaemia. The use of LSTMs is motivated by their remarkable capacity to model the complex temporal dependencies found in data sequences, which makes them particularly suitable for modelling medical time series, where the temporal dynamics of variables is crucial [34]. Furthermore, previous research has demonstrated the effectiveness of LSTM in predicting future BG variability, reinforcing our decision to adopt this architecture to accurately anticipate hypoglycaemic and hyperglycaemic episodes [4, 24]. Formally, an LSTM can be defined as follows:

Let's assume an input sequence $x = (x_1, x_2, \dots, x_t)$, where t stands for time. An LSTM processes this sequence step by step, and at each step t, it takes into account the current input x_t as well as the previous hidden state h_{t-1} and the previous memory cell c_{t-1}. The equations that describe the dynamics of an LSTM are generally formulated as follows [35]:

Forget gate: It determines what information from the previous memory cell c_{t-1} should be forgotten or retained for the new memory cell c_t.

$$f_t = \sigma(w_f.[h_{t-1},\ x_t] + b_f) \tag{3}$$

Input gate: It decides what new information should be added to the memory cell. $\tilde{c}_t$ represents the candidates for the new memory cell.

$$i_t = \sigma(w_i.[h_{t-1}, x_t] + b_i) \tag{4}$$

$$\tilde{c}_t = tanh(w_c.[h_{t-1}, x_t] + b_c) \tag{5}$$

Memory cell update: It is updated based on decisions made by the forget gate and the input gate.

$$c_t = f_t.c_{t-1} + i_t.\tilde{c}_t \tag{6}$$

Output gate: It determines the current output h_t based on the updated memory cell and decides what information should be transmitted to the output. In these equations, σ represents the sigmoid function (logistic function), $tanh$ is the hyperbolic tangent function, W and b are the weights and biases of the network, and [·,·] indicates the concatenation of vectors.

$$o_t = \sigma(w_o.[h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t.\tanh(c_t) \tag{8}$$

In summary, an LSTM employs gates to control the flow of information within the memory cell, enabling it to capture long-term temporal dependencies in data sequences.

4.5 Modelling

Our work stands out due to the utilization of multi-layer LSTM models, allowing for concurrent predictions across 6 PHs at intervals of 5, 10, 15, 20, 25, and 30 minutes. Thus, we adopted an approach where the initial layers are specifically trained for a PH of 30 minutes, thus facilitating the learning of general characteristics inherent in the data, and the last layers are specific to each PH, allowing the model to capture information finer and specialized contextual information. Each set of specific layers receives information specific to the PH it aims to predict. This multi-output LSTM architecture is illustrated in Figure 2.

Our choice of a multi-output LSTM model represents a strategic divergence from the multi-step approach. Indeed, the multi-step approach involves the propagation of prediction errors from one step to the next, which can lead to an accumulation of errors and a degradation of the model's performance on long-term predictions. By opting for a multi-output model, we circumvented this problem by processing all predictions simultaneously, thus avoiding the transmission of errors from one step to another.
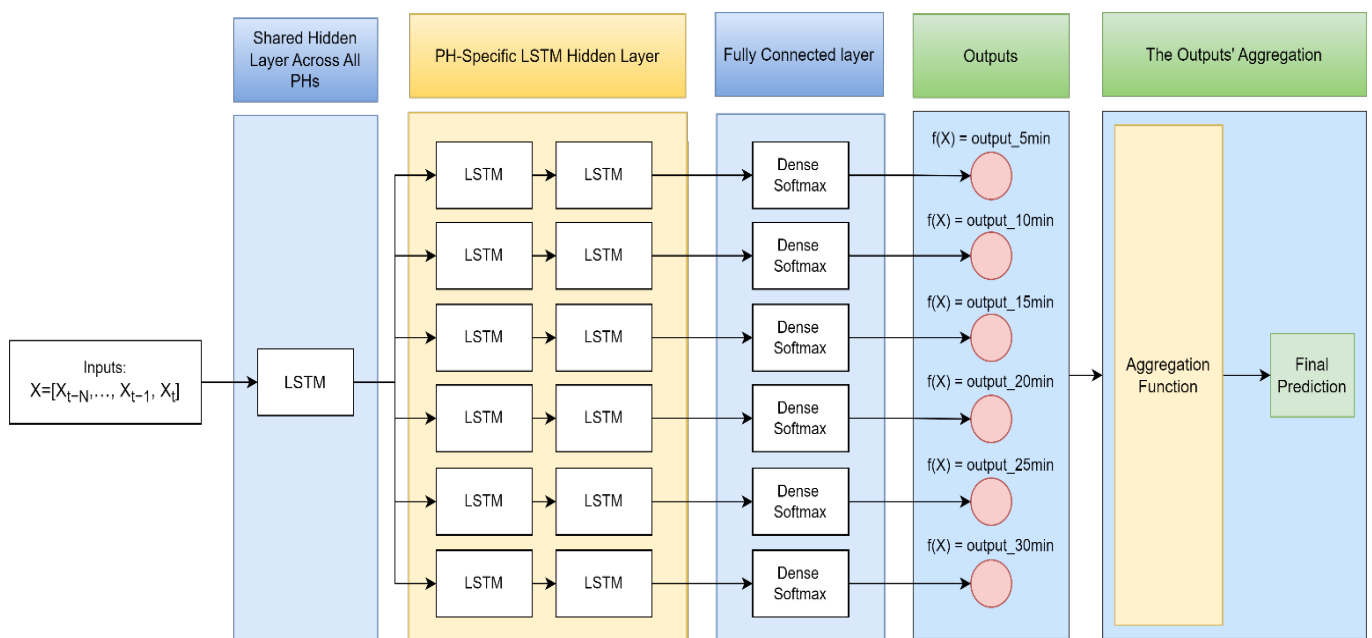

Figure 2. The multi-output LSTM architecture

After the LSTM layers, a dense output layer is added for each specific PHs. Additionally, the SoftMax activation function for binary classification was used to normalize the final outputs, thus providing probabilities associated with each predicted class. Subsequently, a combination approach for the final outputs related to each PH is employed to determine the ultimate prediction, indicating whether an undesirable event will occur. This methodology is detailed in Section 4.7 below.

4.6 Optimization of model parameters

Optimizing a model's predictive performance requires careful selection of hyper-parameters to achieve an optimal balance between model capability and prevention of over fitting.

In this study, our architectural choice was oriented towards the use of an approach where one layer is shared for all PHs, while two specific layers are dedicated to each PH.

The hyper-parameters are adjusted following experiments for a single patient, without resorting to leave-one-out cross-validation. The shared layer was set to 128 LSTM units, while each time horizon-specific LSTM layer is configured with 64 units each. The Adam optimizer [36] was chosen for this study due to its ability to effectively handle the complexities of CGM data, which includes challenges such as BG variability, sensor inaccuracies, noise, and non-stationary patterns. Unlike other optimization algorithms like SGD, RMSProp, and AdaGrad, Adam adapts the learning rates for each parameter based on both the first moment (the mean of the gradients) and the second moment (the variance of the gradients), allowing it to handle noisy and complex datasets efficiently. The learning rate was set to 0.01 initially, but Adam dynamically adjusts it throughout training, ensuring stable convergence and preventing issues like overshooting or local minima. This adaptability makes Adam especially suitable for CGM data. Training spans 50 epochs, with a batch size of 64, and 30% of the data is reserved for validation. The categorical Cross-Entropy loss function is applied to each temporal output. In order to counter over fitting, a dropout layer, with a rate of 20%, is introduced after the shared layer and after the first layer specific to each PH. This approach aims to improve the generalization of the model by randomly deactivating a proportion of the units during the training phase. An early stop mechanism is built in, stopping training if there is no improvement for five consecutive epochs to prevent over fitting.

4.7 Evaluation method

In this work we aim to enhance key performance metrics of the models developed, specifically precision, recall, specificity and F1-score, the definitions of which are provided below.

Precision: measures the proportion of adverse events predicted by the model that are actually predicted. In seeking to increase accuracy, the goal is to minimize FPs, thereby ensuring that the model's positive predictions are highly reliable (see Eq. (9)).

$$\text{Precision} = \frac{\#TPs + \#FPs}{\#TPs} \tag{9}$$

Rappel (Recall): evaluates the model's ability to correctly identify all true adverse events. In seeking to increase recall,

emphasis is placed on minimizing False Negatives (FNs), ensuring that the model does not miss important adverse events (see Eq. (10)).

$$Rappel = \frac{\#TPs + \#FNs}{\#TPs} \tag{10}$$

F1-Score: a metric that considers both precision and recall, providing an overall assessment of model performance. In seeking to improve the F1-score, the goal is to achieve an optimal balance between precision and recall to maximize the overall quality of the model's predictions (see Eq. (11)).

$$\text{F1} - \text{Score} = 2 \times \left(\frac{\text{Precision} + \text{Rappel}}{\text{Precision} \times \text{Rappel}} \right) \tag{11}$$

In these formulas:
True Positives (TPs): these indicate cases where the model correctly predicted an adverse event.
False Positives (FPs): these denote cases where the model incorrectly predicted an adverse event.
False Negatives (TNs): these represent cases where the model failed to predict an adverse event.
True Negative (TNs): these are cases where the model correctly predicts the absence of an adverse event, and this corresponds to reality.
Our model generates six outputs; to obtain an overall prediction over a 30-minute period, we aggregated the predictions for different PHs using indicator notation ($\mathbb{I}$) defined as follow:

$$E = \mathbb{I}\left(\sum_{i}^{n} y_i > 0 \right) \tag{12}$$

where, y_i is the output of the model for the *i-th* PH. The condition states that an event is considered to have occurred if at least one model output is equal to 1. *n* is set to 6 corresponding to 6 PHs.

Subsequently, to measure the effectiveness of our approach to predicting adverse events over a 30-minute period, we compared the list of events that actually occurred to that of predicted events. This comparison was based on the use of the algorithm below which is directly inspired by the previous work [3]. This allows quantifying how close or distant the model predictions are from the actual events.

This algorithm is used to evaluate the performance of an event prediction model by comparing predicted events (P) to actual events (V) over a time scale. The idea is to determine TP, FP, and FN based on a tolerance range around the actual time instants.
Here is the explanation of the algorithm:
  a) **Initialization of sets V and P:**
V is the set of temporal instants associated with real events.
P is the set of predicted events.
  b) **Initializing constants:**
k is a positive constant.
ph is the prediction horizon.
  c) **Initializing result sets:**
TP_set, FP_set, and FN_set are empty sets that will be used to store TPs, FPs, and FNs respectively.
  d) **Iterating through each element tp in P:**
For each element tp in the set of predicted events P, the algorithm searches for a real time instant tv in the set of real events V such that the difference between tp et tv is within the tolerance range defined by -k < tp – tv < ph.

e) Searching for a tv in V:

The algorithm goes through all the elements of the set V to find a tv that satisfies the condition.

f) Processing of results:

If a tv is found, tp is considered a TP and tv is removed from the set V to avoid multiple consideration of the same event. If no tv is found, tp is considered a FP.

g) Identification of FN:

The remaining elements in set V after scanning all predicted events are considered FN.

In the context of BG sensing, prediction distances can be interpreted as the temporal difference between when the model predicted an event and the actual time when that event occurred.

Algorithm

```
V = {t1, ..., tv; # Set of temporal moments associated with
real events
P = {t1, ..., tp; # Set of predicted events
# Initializing constants
k = 3;
ph = 6;
# Initializing result sets
TP_set = set();
FP_set = set();
FN_set = set();
# Iterating through each element tp in P.
for tp in P:
    # Finding a tv in V such that -k < Δtpv < ph
    found_tv = False;
    min_distance = float('inf');
    for tv in V:
        delta_tpv = tp – tv;
        if -k < delta_tpv < ph:
            found_tv = True;
            min_distance       =       min(min_distance,
            abs(delta_tpv));
            matching_tv = tv;
    # If a tv is found, tp is considered a TP
    if found_tv:
        TP_set.add(tp)
        V.remove(matching_tv)  # Remove tv from set V
    # Otherwise, tp is considered a FP
    else:
        FP_set.add(tp)
FN_set = V
```

$$\text{Prediction Distance} = \text{Actual Event Time} - \text{Predicted Event Time} \quad (13)$$

A prediction distance close to zero indicates that the model made a prediction very close to the actual time of the event.

A positive prediction distance suggests that the model anticipated the event, while a negative distance suggests a delay from the actual event.

Below are the different conditions that can be associated with a predicted event.

- $\Delta t_{pv} < -k$: The event is out of analysis range.
- $-k \leq \Delta t_{pv} < 0$: The actual event occurs after the predicted time tp, but still within the tolerance range.
- $\Delta t_{pv} = 0$: The actual event occurs exactly at the predicted time.
- $0 < \Delta t_{pv} < ph$: The actual event tv occurs before the predicted time tp, but still in the future.
- $\Delta t_{pv} \geq ph$: The actual event has already happened. His prediction failed.

5. RESULTS

This section presents the results of our approach, focusing on the performance of the hypoglycaemia and hyperglycaemia prediction models developed from the full dataset, as well as those based on the diurnal and nocturne datasets. Performance metrics, including precision, recall, and F1-score, were calculated for each set and event. The results are presented as mean values and standard deviation.

5.1 Models based on the full dataset

The performances of the hypoglycaemia and hyperglycaemia models, trained on the entire dataset, are presented in Table 2. The models demonstrate appreciable ability to predict hypoglycaemic and hyperglycaemic events, with noteworthy F1-scores of 0.668 and 0.870, respectively.

These results are comparable to the state of the art, indicating significant performance [24]. However, the results reveal significant variability in performance depending on daytime and night-time periods, highlighting the impact of temporality on the predictive capacity of the models. Overall, the prediction of hypoglycaemia shows lower performance compared to the prediction of hyperglycaemia for both periods, highlighting specific challenges in predicting this event.

Diurnal performance shows a decrease in performance compared to night-time prediction of hypoglycaemic events. On the other hand, for the predictions of hyperglycaemia, the results display comparable F1-scores, between the daytime and night-time periods (0.859 and 0.846 respectively), with a higher precision for the diurnal period (0.883) and a higher recall for the night-time period (0.887). Indeed, High precision is important to avoid false alarms, while high recall is crucial to not miss real events.

Table 2. Evaluation results of the model for predicting hypoglycaemia and hyperglycaemia were assessed using the complete dataset, and the performance was measured on the test data from the entire dataset, as well as specifically on the daytime and night-time test data subsets

	Hypoglycaemia				Hyperglycaemia			
	Precision	Recall	Specificity	F1-Score	Precision	Recall	Specificity	F1-Score
All day	0.766 ± 0.071	0.591 ± 0.090	0.979 ± 0.012	0.668 ± 0.057	0.880 ± 0.065	0.860 ± 0.062	0.860 ± 0.039	0.870 ± 0.059
Diurnal	0.716 ± 0.159	0.546 ± 0.070	0.972 ± 0.011	0.619 ± 0.078	0.883 ± 0.054	0.836 ± 0.034	0.821 ± 0.045	0.859 ± 0.033
Nocturn	0.784 ± 0.094	0.594 ± 0.091	0.981 ± 0.010	0.676 ± 0.065	0.808 ± 0.065	0.887 ± 0.053	0.774 ± 0.084	0.846 ± 0.062

Table 3. Evaluation results for hypoglycaemia and hyperglycaemia prediction using distinct daytime and night-time models

	Hypoglycaemia				Hyperglycaemia			
	Precision	Recall	Specificity	F1-Score	Precision	Recall	Specificity	F1-Score
Diurnal	0.729 ±0.181	0.530 ±0.077	0.975 ± 0.007	0.609 ±0.082	0.858 ±0.062	0.874 ±0.044	0.763 ± 0.051	0.866 ±0.0352
Nocturn	0.731 ±0.121	0.628 ±0.095	0.974 ± 0.016	0.675 ±0.056	0.798 ± 0.137	0.882 ± 0.051	0.760 ± 0.087	0.838 ± 0.094

Table 4. Statistics concerning the distances between the predicted events and the actual events for the model developed based on the full dataset, presented in terms of timestamps

	Hypo			Hyper		
	Mean	SD	Median	Mean	SD	Median
Diurnal	-0.231	2.306	-2.0	-1.338	1.609	-2.0
Nocturne	-0.495	2.211	-2.0	-1.628	1.235	-2.0
All day	-0.417	2.236	-2.0	-1.470	1.457	-2.0

Table 5. Statistics concerning the distances between predicted events and real events for the model developed using both the daytime or night-time dataset, presented in terms of timestamps

	Hypo			Hyper		
	Mean	SD	Median	Mean	SD	Median
Diurnal	-0.174	2.329	-2.0	-1.383	1.548	-2.0
Nocturne	-0.553	2.188	-2.0	-1.622	1.258	-2.0

5.2 Models based on diurnal and nocturne datasets

The performances of the models trained specifically on diurnal and nocturne datasets are presented in Table 3.

The specific models dedicated to predicting hypoglycaemia exhibit slightly lower performances than the global model based on the full dataset. However, it is remarkable that these models achieve interesting performances, even if they are developed on a smaller dataset compared to the global models.

Regarding the prediction of hypoglycaemia, the nocturne model shows an improvement in precision (0.731), recall (0.628) and F1-score (0.675) compared to the diurnal model. For the prediction of hyperglycaemia, the diurnal model shows superior performance in terms of precision (0.858) and F1-score (0.866), while the nocturne model shows a slightly higher recall of (0.882).

5.3 Event detection

Based on the list of events predicted over a 30-minute PH, apart from whether it is a TP or a FP, we calculated the statistical measures: median, mean and Standard Deviation (SD), relating to the distance between the detected events and the real events. This means, for each detected event, what is its distance from the real event.

Table 4 presents statistics on the temporal deviations between predicted events (P) and actual events (V), within a PH of 30 minutes. These results come from the model developed on the basis of the full dataset, and are segmented according to the time periods diurnal, nocturne and all day. Additionally, Table 5 presents the statistics relating to the models developed from the diurnal and nocturnal datasets.

The results of the analysis significantly reveal an inclination to anticipate the detection of hypoglycaemic and hyperglycaemic events for global, diurnal and nocturne models. This trend is perceptible through the negative values of the means and medians, thus illustrating a predisposition to identify these events before their actual occurrence. Additionally, the SD highlights the variability inherent in these

distances, indicating that predictions can sometimes deviate from the predicted mean (7±1 minutes for hypoglycaemia and 5±1 minutes for hyperglycaemia). It is relevant to note that the SD associated with hyperglycaemic patterns (2.2±0.1 timestamps) is lower than that of hypoglycaemic events (1.35±0.15 timestamps), suggesting greater consistency in the detection of hyperglycaemic events compared to the variability observed in the detection of hypoglycaemic events. This observation highlights a certain robustness in the models' ability to predict hyperglycaemic events with relatively consistent accuracy.

Medians of -2.0 timestamps, corresponding to 5 minutes, indicate that half of the time deviations are equal to or less than -2.0 timestamps.

6. DISCUSSION

This study assesses the performance of models in predicting hypoglycaemic and hyperglycaemic events, emphasizing the influence of temporal variations on their accuracy. The initial hypothesis suggests that differences in glycaemic behaviour between diurnal and nocturnal periods necessitate a tailored approach, with prediction models developed from distinct datasets specific to each time period.

The performance of models based on the full dataset is slightly better and this can be explained by the fact that they likely benefit from a greater diversity of data, which may contribute to their ability to generalize and predict effectively.

The ability of the specific diurnal and nocturne models to maintain high performance even with smaller datasets suggests some adaptability. This could indicate that specific temporal features can be learned efficiently with smaller datasets, while still maintaining high relevance for prediction.

The relatively lower performance in predicting hypoglycaemia may be attributed to the intrinsic complexity of these events. Hypoglycaemia can occur quickly and requires increased sensitivity of models to detect them early.

The use of RNN, in particular LSTMs, to model the

temporal sequences of glycaemic data is consistent with the dynamic and sequential nature of medical data. LSTMs are well suited to capture complex temporal dependencies, making them valuable in the context of BG, where changes can occur quickly and are strongly influenced by past events. This allows the model to take historical information into account, providing a better understanding of BG trends and improving prediction accuracy.

Using multi-layer LSTM models with simultaneous predictions over multiple PHs provides the flexibility to capture both general data characteristics and contextual information specific to each PH, thereby improving the model's ability to anticipate glycaemic events.

In this study, we opted for an architectural approach where one layer is shared across all prediction horizons (PHs), while two distinct layers are allocated to each individual PH. However, it is important to note that this architectural choice is not exhaustive, and other configurations deserve to be explored regarding the number of shared layers and distinct layers. The complexity of glycaemic dynamics can vary significantly across time horizons, and adjusting the number of layers based on each PH output could potentially improve the model's ability to capture these nuances.

The decision to opt for an evaluation which defines a binary variable indicating the presence (1) or absence (0) of the adverse event in the time interval [t+1,t+PH], contributes to maintain model stability and accuracy across PH of 30 minutes.

The distance Δt_{pv} is a way to quantify the temporal proximity between predicted and actual events, and it is used to make decisions about the accuracy of predictions in a given context. The value of the constants k and ph influences the tolerance granted to this temporal proximity.

The similarity of the results between the models developed based on "Diurnal", "Nocturnal" and "All day" periods highlight their relative stability throughout the day. This temporal consistency suggests that the models maintain their performance regardless of the specific period considered.

It is worth noting that conducting a thorough comparison of model performance faces challenges due to differences in datasets, variations in dataset sizes, distinctions in data pre-processing methods, and disparities in parameters adjustment. Thus, the comparison of the results displayed in the Table 1 lacks accuracy and substantive significance [3]. Consequently, studies relying on the OHIOT1DM dataset offer the most suitable grounds for comparison. Our approach yields highly comparable results with those presented in Table 1 based on the OHIOT1DM dataset.

From a clinical standpoint, the results of this study carry significant practical clinical implications, particularly due to the use of real-world patient data for both model training and evaluation. The BG measurements were captured by a clinically certified CGM sensor, ensuring that the data reflect real-life conditions.

The performance metrics suggest that the model can be used as a decision-support tool for alerting potential adverse events, especially given the promising results achieved [37]. However, further validation with additional real-world datasets is essential to fully confirm its robustness across diverse populations. While the model demonstrates promising results, areas for improvement remain, including reducing false negatives to minimize the risk of missing critical hypoglycaemic or hyperglycaemic events. Exploring other ML approaches, based on night-time and daytime separation, could improve prediction performances, ensuring that the

model is reliable enough for broader real-world clinical use.

The development of predictive models in healthcare, particularly those utilizing CGM data, presents several significant societal concerns that must be addressed. Data privacy and security are paramount, as sensitive CGM information can lead to privacy violations if not properly secured; implementing strong measures like encryption and regulatory compliance is essential. Additionally, algorithmic bias can result in inaccurate predictions for diverse populations if models are trained on unrepresentative datasets, highlighting the need for varied training data and continuous performance monitoring. There is also the risk of over-reliance on AI, which can diminish the essential role of healthcare professionals; thus, AI should be viewed as a supportive tool that requires human oversight. Access to advanced technologies may exacerbate health inequities, particularly in low-income or rural areas, necessitating efforts to enhance access through public healthcare integration and local partnerships. Furthermore, misinterpretation of model predictions can lead to inappropriate medical decisions, making it crucial to provide clear explanations and training for healthcare providers. Lastly, the psychological impacts of AI predictions, such as anxiety from critical alerts, should be mitigated by prioritizing essential notifications and minimizing unnecessary alerts. Proactively tackling these concerns will guarantee the ethical and responsible implementation of BG prediction healthcare applications [37, 38].

7. CONCLUSION

In conclusion, our study explores the problem of prediction of adverse events: hypoglycaemia and hyperglycaemia in T1D patients with a particular intention on the influence of development of models based on temporal data specific to the daytime and night-time periods. Indeed, the developed models, based on the multi-output LSTM architecture, demonstrate remarkable performance comparable to those in the literature. The results reveal that models leveraging the full dataset perform slightly better, suggesting that the increased diversity of the data contributes to their effective generalization and prediction ability. It is interesting to note that despite a more restricted data set, the models dedicated to daytime and night-time periods maintain excellent performance. This constancy highlights the adaptability of the models, suggesting that specific temporal features can be learned efficiently with more limited datasets while maintaining high relevance for prediction. In sum, our study highlights the crucial importance of taking into account temporal variations in the development of glycaemic prediction models. The results obtained open interesting perspectives for improving the specificity and sensitivity of the models, paving the way for future developments in the prediction and detection of adverse events, based on specific temporal periods, namely the diurnal period and the nocturnal period.

REFERENCES

[1] Standl, E., Khunti, K., Hansen, T.B., Schnell, O. (2019). The global epidemics of diabetes in the 21st century: Current situation and perspectives. European Journal of Preventive Cardiology, 26(2): 7-14.

https://doi.org/10.1177/2047487319881021

[2] Seo, W., Lee, Y.B., Lee, S., Jin, S.M., Park, S.M. (2019). A machine-learning approach to predict postprandial hypoglycemia. BMC Medical Informatics and Decision Making, 19: 1-13. https://doi.org/10.1186/s12911-019-0943-4

[3] Gadaleta, M., Facchinetti, A., Grisan, E., Rossi, M. (2018). Prediction of adverse glycemic events from continuous glucose monitoring signal. IEEE Journal of Biomedical and Health Informatics, 23(2): 650-659. https://doi.org/10.1109/JBHI.2018.2823763

[4] Yang, M., Dave, D., Erraguntla, M., Cote, G.L., Gutierrez-Osuna, R. (2022). Joint hypoglycemia prediction and glucose forecasting via deep multi-task learning. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, pp. 1136-1140. https://doi.org/10.1109/ICASSP43922.2022.9746129

[5] Marling, C., Bunescu, R. (2020). The OhioT1DM dataset for blood glucose level prediction: Update 2020. In CEUR Workshop Proceedings, NIH Public Access, pp. 71-74.

[6] Dubosson, F., Ranvier, J.E., Bromuri, S., Calbimonte, J.P., Ruiz, J., Schumacher, M. (2018). The open D1NAMO dataset: A multi-modal dataset for research on non-invasive type 1 diabetes management. Informatics in Medicine Unlocked, 13: 92-100. https://doi.org/10.1016/j.imu.2018.09.003

[7] Mayo, M., Koutny, T. (2020). Neural multi-class classification approach to blood glucose level forecasting with prediction uncertainty visualisation. In KDH 2020, CEUR Workshop Proceedings, pp. 80-84.

[8] Bertachi, A., Biagi, L., Contreras, I., Luo, N., Vehí, J. (2018). Prediction of blood glucose levels and nocturnal hypoglycemia using physiological models and artificial neural networks. In KDH@ IJCAI, pp. 85-90.

[9] Aliberti, A., Pupillo, I., Terna, S., Macii, E., Di Cataldo, S., Patti, E., Acquaviva, A. (2019). A multi-patient data-driven approach to blood glucose prediction. IEEE Access, 7: 69311-69325. https://doi.org/10.1109/ACCESS.2019.2919184

[10] Vu, L., Kefayati, S., Idé, T., Pavuluri, V., Jackson, G., Latts, L., Chang, Y.C. (2020). Predicting nocturnal hypoglycemia from continuous glucose monitoring data with extended prediction horizon. In AMIA Annual Symposium Proceedings, pp. 874-882.

[11] Güemes, A., Cappon, G., Hernandez, B., Reddy, M., Oliver, N., Georgiou, P., Herrero, P. (2019). Predicting quality of overnight glycaemic control in type 1 diabetes using binary classifiers. IEEE Journal of Biomedical and Health Informatics, 24(5): 1439-1446. https://doi.org/10.1109/JBHI.2019.2938305

[12] Afentakis, I., Unsworth, R., Herrero, P., Oliver, N., Reddy, M., Georgiou, P. (2023). Development and validation of binary classifiers to predict nocturnal hypoglycemia in adults with type 1 diabetes. Journal of Diabetes Science and Technology, 19(1): 153-160. https://doi.org/10.1177/19322968231185796

[13] Zhang, M., Flores, K.B., Tran, H.T. (2021). Deep learning and regression approaches to forecasting blood glucose levels for type 1 diabetes. Biomedical Signal Processing and Control, 69: 102923. https://doi.org/10.1016/j.bspc.2021.102923

[14] Bertachi, A., Viñals, C., Biagi, L., Contreras, I., Vehí, J.,
Conget, I., Giménez, M. (2020). Prediction of nocturnal hypoglycemia in adults with type 1 diabetes under multiple daily injections using continuous glucose monitoring and physical activity monitor. Sensors, 20(6): 1705. https://doi.org/10.3390/s20061705

[15] Jensen, M.H., Christensen, T.F., Tarnow, L., Seto, E., Dencker Johansen, M., Hejlesen, O.K. (2013). Real-time hypoglycemia detection from continuous glucose monitoring data of subjects with type 1 diabetes. Diabetes Technology & Therapeutics, 15(7): 538-543. https://doi.org/10.1089/dia.2013.0069

[16] Oviedo, S., Contreras, I., Quirós, C., Giménez, M., Conget, I., Vehi, J. (2019). Risk-based postprandial hypoglycemia forecasting using supervised learning. International Journal of Medical Informatics, 126: 1-8. https://doi.org/10.1016/j.ijmedinf.2019.03.008

[17] Fox, I., Ang, L., Jaiswal, M., Pop-Busui, R., Wiens, J. (2018). Deep multi-output forecasting: Learning to accurately predict blood glucose trajectories. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, United Kingdom, pp. 1387-1395. https://doi.org/10.1145/3219819.3220102

[18] Li, K., Daniels, J., Liu, C., Herrero, P., Georgiou, P. (2019). Convolutional recurrent neural networks for glucose prediction. IEEE Journal of Biomedical and Health Informatics, 24(2): 603-613. https://doi.org/10.1109/JBHI.2019.2908488.

[19] Daniels, J., Herrero, P., Georgiou, P. (2021). A multitask learning approach to personalized blood glucose prediction. IEEE Journal of Biomedical and Health Informatics, 26(1): 436-445. https://doi.org/10.1109/JBHI.2021.3100558

[20] Dudukcu, H.V., Taskiran, M., Yildirim, T. (2021). Blood glucose prediction with deep neural networks using weighted decision level fusion. Biocybernetics and Biomedical Engineering, 41(3): 1208-1223. https://doi.org/10.1016/j.bbe.2021.08.007

[21] Şahin, A., Aydın, A. (2021). Personalized advanced time blood glucose level prediction. Arabian Journal for Science and Engineering, 46(10): 9333-9344. https://doi.org/10.1007/s13369-020-05263-2

[22] Alfian, G., Syafrudin, M., Anshari, M., Benes, F., Atmaji, F.T.D., Fahrurrozi, I., Rhee, J. (2020). Blood glucose prediction model for type 1 diabetes based on artificial neural network with time-domain features. Biocybernetics and Biomedical Engineering, 40(4): 1586-1599. https://doi.org/10.1016/j.bbe.2020.10.004

[23] Zulj, S., Carvalho, P., Ribeiro, R.T., Andrade, R., Magjarevic, R. (2021). Data size considerations and hyperparameter choices in case-based reasoning approach to glucose prediction. Biocybernetics and Biomedical Engineering, 41(2): 733-745. https://doi.org/10.1016/j.bbe.2021.04.013

[24] D'Antoni, F., Petrosino, L., Marchetti, A., Bacco, L., Pieralice, S., Vollero, L., Merone, M. (2023). Layered meta-learning algorithm for predicting adverse events in type 1 diabetes. IEEE Access, 11: 9074-9094. https://doi.org/10.1109/ACCESS.2023.3237992

[25] Nemat, H., Khadem, H., Eissa, M.R., Elliott, J., Benaissa, M. (2022). Blood glucose level prediction: Advanced deep-ensemble learning approach. IEEE Journal of Biomedical and Health Informatics, 26(6): 2758-2769. https://doi.org/10.1109/JBHI.2022.3144870

[26] Georga, E.I., Protopappas, V.C., Ardigo, D., Polyzos, D., Fotiadis, D.I. (2013). A glucose model based on support vector regression for the prediction of hypoglycemic events under free-living conditions. Diabetes Technology & Therapeutics, 15(8): 634-643. https://doi.org/10.1089/dia.2012.0285

[27] Prendin, F., Del Favero, S., Vettoretti, M., Sparacino, G., Facchinetti, A. (2021). Forecasting of glucose levels and hypoglycemic events: Head-to-head comparison of linear and nonlinear data-driven algorithms based on continuous glucose monitoring data only. Sensors, 21(5): 1647. https://doi.org/10.3390/s21051647

[28] Jung, M., Lee, Y.B., Jin, S.M., Park, S.M. (2017). Prediction of daytime hypoglycemic events using continuous glucose monitoring data and classification technique. arXiv preprint arXiv:1704.08769. https://doi.org/10.48550/arXiv.1704.08769

[29] Nguyen, L.L., Su, S., Nguyen, H.T. (2014). Neural network approach for non-invasive detection of hyperglycemia using electrocardiographic signals. In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, pp. 4475-4478. https://doi.org/10.1109/EMBC.2014.6944617

[30] Jensen, M.H., Dethlefsen, C., Vestergaard, P., Hejlesen, O. (2020). 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, 14(2): 250-256. https://doi.org/10.1177/1932296819868727

[31] Felizardo, V., Garcia, N.M., Megdiche, I., Pombo, N., Sousa, M., Babič, F. (2023). Hypoglycaemia prediction using information fusion and classifiers consensus. Engineering Applications of Artificial Intelligence, 123: 106194. https://doi.org/10.1016/j.engappai.2023.106194

[32] American Diabetes Association. (2021). 6. Glycemic targets: Standards of medical care in diabetes—2021. Diabetes Care, 44(1): S73-S84. https://doi.org/10.2337/dc21-S006

[33] Camerlingo, N., Vettoretti, M., Del Favero, S., Cappon, G., Sparacino, G., Facchinetti, A. (2019). A real-time continuous glucose monitoring–based algorithm to trigger hypotreatments to prevent/mitigate hypoglycemic events. Diabetes Technology & Therapeutics, 21(11): 644-655. https://doi.org/10.1089/dia.2019.0139

[34] Yu, Y., Si, X., Hu, C., Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. Neural Computation, 31(7): 1235-1270. https://doi.org/10.1162/neco_a_01199

[35] Hochreiter, S., Long Short-term Memory. Neural Computation MIT-Press, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[36] Kinga, D., Adam, J.B. (2015). A method for stochastic optimization. In International Conference on Learning Representations (ICLR), USA, p. 6.

[37] Papazafiropoulou, A.K. (2024). Diabetes management in the era of artificial intelligence. Archives of medical sciences. Atherosclerotic Diseases, 9: e122. https://doi.org/10.5114/amsad/183420

[38] Singla, R., Singla, A., Gupta, Y., Kalra, S. (2019). Artificial intelligence/machine learning in diabetes care. Indian Journal of Endocrinology and Metabolism, 23(4): 495-497. https://doi.org/10.4103/ijem.ijem_228_19