2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340094

Jointly Predicting Postprandial Hypoglycemia and Hyperglycemia Using Continuous Glucose Monitoring Data in Type 1 Diabetes

Ran Cui[1], Christopher J Nolan[2], Elena Daskalaki[1], and Hanna Suominen[1,2,3]

***Abstract*— The development of continuous glucose monitoring (CGM) systems has enabled people with type 1 diabetes mellitus (T1DM) to track their glucose trajectory in real-time and inspired research in personalised glucose prediction. In this paper, our aim is to predict postprandial abnormal-glycemia events. Different from prior research which focuses on hypoglycemia only, we make the first attempt to establish our problem as the joint prediction of hyperglycemia and hypoglycemia. On this basis, we propose a machine learning model that learns from the pattern of 1 hour past glucose and makes predictions for the two tasks simultaneously using a unified backbone. Key benefits of our methodology include 1) requiring only the CGM sequence as the input, thus making it more widely applicable than other counterparts using extra inputs such as the nutrition details, and 2) minimising the computational cost as the two tasks are unified into a single model. Our experiments on the openly available OhioT1DM dataset achieve state-of-the-art performance (Matthew's correlation coefficient of 0.61 for hyperglycemia and 0.48 for hypoglycemia). To encourage further study, we release our codes at `https://github.com/r-cui/PostprandialHyperHypoPrediction` under the MIT license.**

I. Introduction

Type 1 diabetes mellitus (T1DM) is a chronic metabolic disorder affecting millions of people globally. Many people with T1DM wear continuous glucose monitoring (CGM) devices to constantly track their glucose trajectory [1], and personalised glucose prediction using machine learning (ML) on CGM data has drawn significant research attention [2], [3].

To date, the most straightforward setting of glucose prediction is short-term glucose prediction [4]–[9]. In this setting, a model is trained and evaluated using the full set of CGM data, and the learning target is to accurately predict the glucose value at a fixed horizon such as 30 or 60 minutes using recent glucose-related information as the input. However, glucose presents different patterns during the day. As shown in Figure 1, under the effect of food, postprandial glucose usually has a pattern from rising to declining that differs from other times such as the nocturnal glucose, which is relatively stable. Table I shows a statistical comparison of postprandial 4-hour glucose with glucose at other times using the OhioT1DM dataset [10]. The two-sample Kolmogorov–Smirnov (KS) tests [11] on all patients achieve a p-value less than 0.05, indicating their difference in probability distribution. This evidence suggests the potential of glucose prediction in a finer granularity, and, in this study, we specifically target the postprandial scenario.

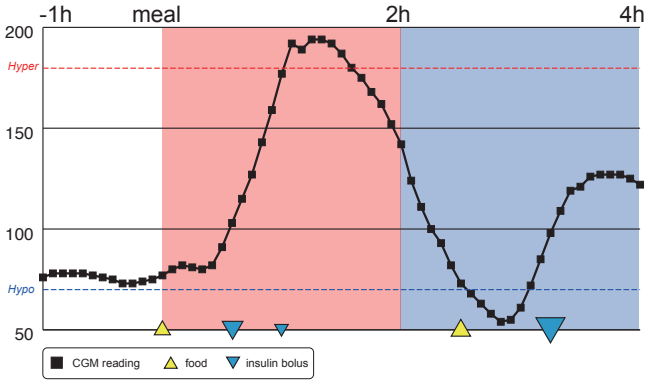

Fig. 1. A snippet of T1DM patient's real-life CGM data around a meal. This example shows a typical postprandial glucose pattern: the digestion of food causes glucose to peak, followed by its falling to a valley under the effect of natural metabolism and meal-corresponded insulin. Consequently, hyperglycemia and hypoglycemia are both risky in postprandial times.

Prior research has mainly approached postprandial glucose prediction from two directions, namely prediction at meal time [12]–[15] and prediction in fixed horizon [16], [17]. Meal-time prediction considers a pre-defined meal effective time range such as 4 hours after the meal, and all predictions regarding the time range are made at once at the meal time. This scheme reflects the application that the patient could re-plan the meal advised by the prediction result, e.g., adjust the planned food quantity or insulin amount. Although the envisaged application is of great value, a natural disadvantage of this scheme is that it cannot take factors after the meal intake into consideration. As illustrated in Figure 1, other meals and insulin intake events could be present at the patient's wish in real life, rendering the meal-time prediction inaccurate under such circumstances. In contrast, the fixed prediction horizon scheme considers a shorter prediction horizon such as 30 or 60 minutes, and the prediction is rolling starting right after the meal time until it covers all the meal effective time range. This scheme corresponds to the application that once a meal event is announced by the patient, the algorithm constantly runs to predict glucose for a short future, such that later events that reflect on glucose profile can be accommodated. Although having a shorter prediction horizon than the meal-time prediction scheme, the application of fixed horizon prediction is more flexible.

[1]School of Computing, The Australian National University, Canberra, Australia
[2]School of Medicine and Psychology, The Australian National University, Canberra, Australia
[3]Department of Computing, University of Turku, Turku, Finland
ran.cui, christopher.nolan, eleni.daskalaki, hanna.suominen@anu.edu.au

TABLE I

DISTRIBUTION IN THE FORMAT OF MEAN(STD) OF POSTPRANDIAL 4-HOUR CGM DATA AND OTHER CGM DATA IN THE OHIOT1DM DATASET, AND TWO-SAMPLE KOLMOGOROV–SMIRNOV (KS) TESTS PER PATIENT.

Patient ID	540	544	552	584	596	559	563	570	575	588	591
Postprandial 4h CGM	139.7 (60.7)	183.5 (59.5)	145.3 (55.7)	186.1 (64.2)	147.9 (51.1)	173.8 (72.8)	149.8 (51.3)	181.1 (66.2)	137.3 (62.3)	163.5 (55.3)	151.2 (57.9)
Other CGM	136.0 (53.2)	142.5 (52.7)	147.2 (54.0)	195.4 (65.7)	142.7 (46.0)	155.1 (66.6)	142.8 (48.3)	192.8 (56.7)	142.0 (56.9)	164.2 (47.7)	158.1 (58.2)
KS Test $p < 0.05$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

In this study, we aim to design and develop a strategy for the prediction of postprandial abnormal-glycemia events under the fixed horizon prediction scheme. Specifically, we propose a problem formulation under the fixed prediction horizon scheme in which the prediction target is set to both postprandial hyperglycemia and hypoglycemia instead of only one of them as in previous studies [12], [13], [17]. The two targets are annotated simply by two distinctive configurations of time zones and thresholds (Figure 1) according to clinical suggestions [18], [19], so this expansion does not add extra complexity to the problem. Under our problem formulation, we propose a long short-term memory (LSTM) [20] based model that uses the 1-hour past CGM pattern to learn to predict the near future. Instead of separately obtaining two models for the two tasks (hypo-, hyper-glycemia), our method uses a single LSTM backbone to handle the temporal feature extraction for both tasks, thus the need for computational resources in both training and inference is minimised. As the envisaged user scenario is that the user notifies the system of entering the postprandial stage after each meal and the system constantly runs the prediction, we consider this optimisation important for running on mobile or low-power devices. As a reflection of the recent initiative of data sharing in this field [21], we evaluate our proposed methodology on the openly available OhioT1DM dataset [10] collected in everyday settings. Though there exist differences in the data used by different studies, our experiments achieve state-of-the-art results compared to existing literature. Moreover, to further study the effect of the aforementioned glucose distribution shift in postprandial scenario, we conduct an ablation experiment in which we train the model on the whole CGM dataset. The resulting performance deterioration and 10-fold increase of training process indicate that treating postprandial glucose prediction separately is indeed a valid approach that should be further pursued.

The contributions of our study are summarised as follows:

- To the best of our knowledge, we are the first to formulate the problem of jointly predicting postprandial hyperglycemia and hypoglycemia in the field of postprandial glucose prediction.
- We propose a ML approach that unifies the two prediction tasks into one single model so that the need for computational resources is minimised. It achieves state-of-the-art prediction quality compared to existing studies.
- Our statistical analysis and ablation study verify the impact of the glucose distribution shift problem in postprandial glucose prediction.

II. RELATED WORKS

Our study sits in the cross-domain of *short-term glucose prediction* and *postprandial glucose prediction*. Thus, we review related existing research in both fields.[1]

A. Short-Term Glucose Prediction

We refer to *short-term glucose prediction* [4]–[9] as the research field of predicting the glucose value at a fixed prediction horizon ranging from 15 to 120 minutes using recent glucose-related information as input, including CGM recordings, carbohydrates intake and insulin delivery. To date, most of these studies were evaluated on openly available datasets such as OhioT1DM [10] in purpose of finding better methodology enabled by open performance comparison. Deep learning models have been recently shown to be widely effective on this problem, such as recurrent neural networks (RNN) [9], dilated convolution neural networks (CNN) [7], multi-scale LSTM [8], and self-attention [6].

However, research on this topic was mostly designed and evaluated using all available CGM data. As discussed in the introduction, the unique distribution and pattern of postprandial glucose suggest the potential of distinguishing different scenarios in a day which is inadequately investigated in this field.

B. Postprandial Glucose Prediction

We refer to *postprandial glucose prediction* [12]–[17] as the research field of glucose prediction or hyperglycemia (hypoglycemia) prediction that targets only the postprandial scenario, using either only CGM recording or combine with meal information such as the nutrition components. Research in this field can be classified into two categories differing in whether the prediction is made at the meal time at once or is constantly made with a fixed prediction horizon. For the first category, [12], [13] studied meal-time prediction of postprandial hypoglycemia in 4 hours after the meal and their state-of-the-art performance was 0.48 in Matthew's correlation coefficient. Different in the prediction target, [14] studied meal time prediction of the full excursion of 2 hours postprandial glucose, and [15] studied the meal-time prediction of several features such as the area under curve of 2 hours postprandial and glucose value at 1 hour postprandial that could reveal the future trend of postprandial glucose. However, a natural inadequacy of meal-time prediction is that it is unable to accommodate a dynamic future, such as another meal or extra insulin intake in the meal effective time range as the patient may wish. On the other hand, the second category in

[1]For more comprehensive reviews of glucose prediction, we refer the readers to [2], [3].

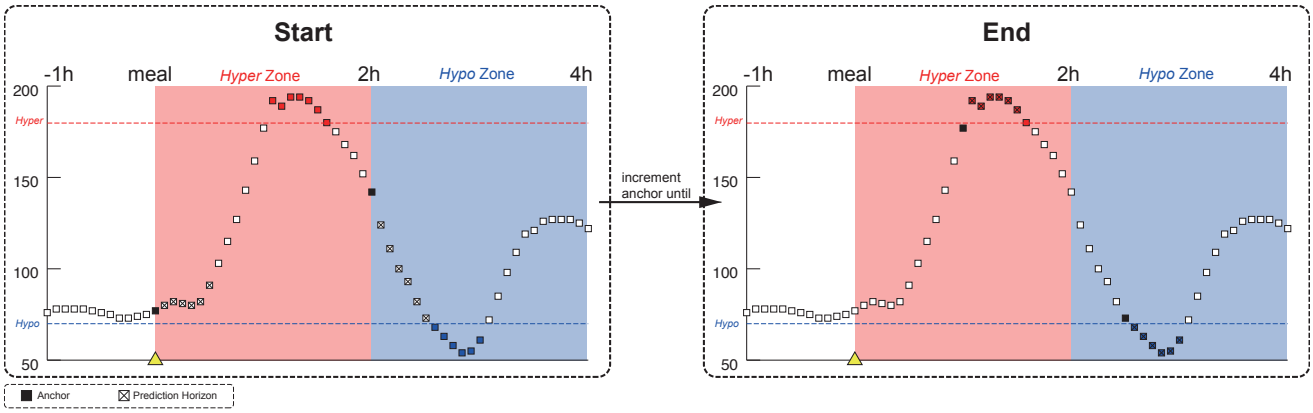

Fig. 2. The illustration of our examples extraction. Taking hyperglycemia for illustration, an anchor started to increment from the first timestamp in the red zone, and a corresponding example was generated at each incrementing step until it arrived just before the occurrence of a hyperglycemia event. The extraction of hypoglycemia was of a similar manner as illustrated in the blue zone.

this field follows the constant prediction strategy using fixed prediction horizon. In this category, [16] studied the use of nutrition absorption models in predicting the glucose value at 1 hour. [17] explored prediction of postprandial hypoglycemia in 30 minutes using random forest classification model and achieved 0.9 sensitivity and 0.9 specificity with a cost of 0.7 false alarm rate.

Our study is in relation to short-term glucose prediction as we follow the fixed prediction horizon strategy, but with a focus on postprandial data only. Moreover, our study is in line with the second category of postprandial glucose prediction, but targeting the joint prediction of postprandial hyperglycemia and hypoglycemia, and relying on only the CGM sequence which circumvents extra burdens such as estimating meal carbohydrates [22], [23].

III. Proposed Method

We present our formulation of postprandial hyperglycemia and hypoglycemia prediction in Section III-A, and the rest of Section III constitutes the description of our ML methodology for this problem.

A. Problem Formulation

Under the configuration of the sampling rate being 5 minutes, we set our prediction horizon to be 30 minutes (6 data points). In accordance with clinical consensus [18], [19], we consider 0-2 hours and 2-4 hours after the meal to be the effective time range for the two tasks of postprandial hyperglycemia and hypoglycemia, respectively. Under this configuration, the 4-hour postprandial glucose trajectory, denoted by $\mathbf{g} = [g_1, g_2, ..., g_{48}]$, was our source of examples extraction for a meal at the time of sample g_0. Specifically, for each task, we applied a moving anchor i starting at the first index of its time range. A new example (i, y_i) was generated each time the anchor incremented until the prediction horizon moved to the last 6 data points in the time range. To this end, the anchor i took a default range in

$$\text{hyperglycemia} \qquad \{i \in \mathbb{N} : 1 \leq i \leq 18\}, \tag{1}$$

$$\text{hypoglycemia} \qquad \{i \in \mathbb{N} : 25 \leq i \leq 42\}. \tag{2}$$

We defined that a hyperglycemia (hypoglycemia) event occurred if at least two consecutive data points in the prediction horizon crossed the threshold of 180 mg/dL (70 mg/dL) [24]. Consequently, the label y_i was a binary value determining by if such event occurred in the prediction horizon $\mathbf{g}_{i+1:i+6}$.

On this basis, an early stopping criterion in the incremental

Algorithm 1 Pseudo Codes for Extracting Train/Test Examples from a Meal.

```
1:  Input: g, t                                   ▷ CGM sequence, meal time
2:  Initialise ph = 6                             ▷ prediction horizon: 0.5 h
3:
4:  Case 1: hyperglycemia
5:  Initialise thres = 180
6:  Initialise T = [1,2,...,18]                   ▷ 0-2 h
7:  procedure ExtractData(g, t)
8:      data = []
9:      for i in T do
10:         if g_{t+i} ≥ thres then
11:             break
12:         end if
13:         if two consecutive in g_{(t+i+1):(t+i+ph)} ≥ thres then
14:             y = True
15:         else
16:             y = False
17:         end if
18:         push tuple (t, i, y) to data
19:     end for
20:     return data
21: end procedure
22:
23: Case 2: hypoglycemia
24: Initialise thres = 70
25: Initialise T = [25,26,...,42]                 ▷ 2-4 h
26: procedure ExtractData(g, t)
27:     Change ≥ to ≤ in line 10 and 13.
28: end procedure
```

process of the anchor i was applied: the increment of i early stopped at the last point before an event of interest being present (Figure 2). This was to reflect that the prediction corresponding to a certain meal was no longer required once the meal already caused a hyperglycemia (hypoglycemia). By repeating the procedure described above on all meals, the full postprandial dataset could be constructed using the raw personalised CGM data. Such dataset was in line with the fixed-horizon prediction research as reviewed in Section II-B. A pseudo code snippet for this training and testing examples generation is shown in Algorithm 1.

B. Classification to Regression

Under this problem formulation, a central idea of our methodology was to change the natural binary classification setup adopted by previous studies [12], [13], [17] into an approximated regression scheme. More specifically, instead of directly making a prediction of the binary y_i, we let our model learn to predict a numerical value of the max (min) glucose value in the prediction horizon, then the binary decision was made via the threshold 180 mg/dL (70 mg/dL). Although this translation was not exactly equivalent to the definition of two-consecutive points crossing the threshold, we adopt this scheme for its stronger supervision to the model training, because the binary supervision would lose the information of the severity of the hyperglycemia (hypoglycemia) event. Moreover, as the events tend to be rare or even do not present at all (especially for hypoglycemia), the example distribution tends to be significantly imbalanced. Our regression scheme also naturally avoided the need of dealing with the imbalanced class problem, such as choosing a proper class weighting in classification loss.

C. Unified Model and Joint Training for Hyperglycemia and Hypoglycemia

As the definition of hyperglycemia and hypoglycemia only differed in the threshold and time range, we approached the two tasks simultaneously using a unified model structure which took the recent glucose pattern as input. To be specific, we used the 1-hour recent CGM sequence $\mathbf{g}_{i-11:i}$ as the input and left out all other factors such as the meal and insulin amount in consideration of the difficulty of obtaining reliable meal amount estimation [25] and that the data may not be sufficiently large to capture the personalised effect of insulin. To encode the sequential feature, we fed the input 1-hour CGM sequence into an LSTM [20] backbone with hidden dimension D followed by a Rectified Linear Unit (ReLU) activation function [26]. The last hidden state $\mathbf{h}$ of the LSTM, given by

$$\mathbf{h} = \max(\text{LSTM}(\mathbf{g}_{i-11:i}), 0) \in \mathbb{R}^D, \tag{3}$$

was regarded as the encoded feature vector of the input CGM sequence. The LSTM backbone was followed by two linear heads, which took the identical $\mathbf{h}$ as the input, but were respectively responsible for the prediction of hyperglycemia and hypoglycemia. To reflect our regression scheme described in Section III-B, the two linear heads projected $\mathbf{h}$ to two numerical outputs that respectively corresponded to the maximum and minimum glucose in the prediction horizon. The two predictions were given by

$$\hat{g}_{\text{hyper}} = \mathbf{W}_{\text{hyper}} \cdot \mathbf{h} + b_{\text{hyper}} \in \mathbb{R}, \tag{4}$$

$$\hat{g}_{\text{hypo}} = \mathbf{W}_{\text{hypo}} \cdot \mathbf{h} + b_{\text{hypo}} \in \mathbb{R}, \tag{5}$$

where $\mathbf{W} \in \mathbb{R}^D$ and $b \in \mathbb{R}$ represented the canonical linear weight and bias term in a linear projection, respectively.

Finally, mean squared error (MSE) loss between the predicted value and the ground truth was employed in supervising the model training, given by

$$\mathscr{L}_{\text{hyper}} = \frac{1}{M}\sum^{M}(\hat{g}_{\text{hyper}} - \max \mathbf{g}_{i+1:i+6})^2, \tag{6}$$

$$\mathscr{L}_{\text{hypo}} = \frac{1}{N}\sum^{N}(\hat{g}_{\text{hypo}} - \min \mathbf{g}_{i+1:i+6})^2, \tag{7}$$

where M and N denote the total number of training examples extracted for hyperglycemia and hypoglycemia, respectively.

As the two linear heads shared the same hidden state $\mathbf{h}$ as the input, the LSTM was encouraged to learn the temporal trend of the recent glucose without bias towards either task, such as predicting a high glucose value for hyperglycemia or a low glucose value for hypoglycemia. To promote this in the training process, in each training iteration, a hyperglycemia batch and a hypoglycemia batch were independently fed to the model, and the model was alternatively updated using the corresponding loss (i.e., $\mathscr{L}_{\text{hyper}}$ or $\mathscr{L}_{\text{hypo}}$, depending on the task of that batch). As will be shown in our ablation study in Section IV-E, our jointly designed methodology for the two tasks hyperglycemia and hypoglycemia achieved comparable performance to the default option of independently training two models for the two tasks, yet half of the computational resources was saved.

IV. Experiments

A. Dataset

We evaluated our proposed method for predicting postprandial hyperglycemia and hypoglycemia using the publicly available dataset OhioT1DM [10] created by the US Ohio University. This dataset contained eight weeks of data from CGM devices, self-reported insulin intake information and meal events for 12 T1DM patients. We made use of the CGM recordings and followed the train/test split officially defined by the author of the dataset in evaluating our method, in which the test set comprised around ten days of data for each patient[2]. Ethical approval (2020/411) was obtained from the Human Research Ethics Committee of The Australian National University to use de-identified data for this study.

B. Evaluation Method

In evaluating the performance of our proposed methodology, we evaluated the root mean square error (RMSE) for our regression output in accordance with existing studies in general glucose prediction, and further evaluated the sensitivity (SE),

[2]In the raw dataset, the patient with ID 567 provided no meal information in the test split, so we excluded this patient from our experiments.

TABLE II

MAIN PERFORMANCE COMPARISON IN THE FORMAT OF MEAN (STD) OVER 10 REPEATED EXPERIMENTS WITH DIFFERENT RANDOM SEEDS. BOLD SCORES WERE STATISTICALLY SIGNIFICANTLY BETTER ($p < 0.05$) THAN ALL BASELINES VIA THE UNPAIRED T TEST.

	Hyperglycemia						Hypoglycemia					
	RMSE↓	SE↑	SP↑	FA↓	MCC↑	DT↑	RMSE↓	SE↑	SP↑	FA↓	MCC↑	DT↑
Short-Term - [4]	19.08	-	-	-	-	-	19.08	-	-	-	-	-
Short-Term - [5]	18.93	-	-	-	-	-	18.93	-	-	-	-	-
Short-Term - [6]	17.97	-	-	-	-	-	17.97	-	-	-	-	-
Postprandial - [17]	-	-	-	-	-	-	-	0.90 (0.03)	0.91 (0.02)	0.70 (0.04)	-	25.5 (1.97)
Postprandial - [13]	-	-	-	-	-	-	-	0.69	0.8	-	0.24	-
Postprandial - [12]	-	-	-	-	-	-	-	0.71	0.79	-	0.48	-
[17] Replicated	-	0.83 (0.01)	0.81 (0.00)	0.58 (0.01)	0.50 (0.00)	19.28 (0.23)	-	0.82 (0.00)	0.94 (0.00)	0.85 (0.00)	0.34 (0.01)	19.4 (0.00)
Baseline - Dummy	27.01	0.03	1.00	0.00	0.15	5.00	17.44	0.05	1.00	0.40	0.17	5.00
Baseline - AdaBoost	22.03 (0.07)	0.71 (0.01)	0.88 (0.01)	0.46 (0.01)	0.53 (0.01)	16.53 (0.28)	17.42 (0.14)	0.07 (0.02)	1.00 (0.00)	0.70 (0.09)	0.13 (0.03)	10.25 (2.36)
Baseline - Random Forest	19.45 (0.04)	0.58 (0.01)	0.94 (0.00)	0.33 (0.01)	0.55 (0.01)	14.65 (0.19)	14.14 (0.02)	0.31 (0.01)	1.00 (0.00)	0.48 (0.01)	0.39 (0.01)	8.93 (0.23)
Baseline - MLP	18.79 (0.21)	0.57 (0.01)	0.94 (0.00)	0.33 (0.01)	0.55 (0.01)	14.79 (0.28)	13.23 (0.19)	0.38 (0.05)	0.99 (0.00)	0.51 (0.04)	0.42 (0.02)	8.94 (1.54)
Ours	**18.23** (0.35)	0.66 (0.04)	0.95 (0.01)	0.33 (0.03)	**0.61** (0.02)	15.01 (0.98)	13.25 (0.17)	0.48 (0.05)	0.99 (0.01)	0.50 (0.05)	**0.48** (0.02)	10.97 (0.89)

specificity (SP), false alarm rate (FA) which were calculated out of the confusion matrix for the final classification. We also included Matthew's correlation coefficient (MCC) as an overall reflection of the classification result. Additionally, to evaluate the timeliness of our prediction, we calculated the detection time (DT) of the start of a hyperglycemia or hypoglycemia event averaged over the meals that were correctly predicted by our algorithm. In these metrics, MCC took values in $[-1,1]$ to reflect the overall classification quality and others in $[0,1]$, larger (smaller) values reflected better performance for SE, SP, and MCC (FA). For DT, the theoretically optimal value was 30 minutes as it was our preset prediction horizon.

C. Implementation Details

Due to the small size of the patient cohort, we trained personalised model for each patient in our experiments. The full training and evaluation were implemented using PyTorch. Normalisation was applied to all CGM data using mean and standard deviation estimated via training data. In the model configuration, we applied a uni-layer LSTM with hidden dimension $D = 50$, resulting in the full model containing around 10k trainable parameters. The model was tuned using Adam optimiser [27] with batch size 64 and learning rate of 0.001 under a gradient norm clipping at 1.0. To illustrate the computational efficiency under this configuration, each training step took around 0.01 seconds, and the full training pipeline for all 11 patients in OhioT1DM dataset took around 4 minutes on an Apple M1 CPU. As the amount of personalised postprandial CGM data was limited, instead of using hold-out training data as a validation set and conduct model selection, we trained our model on the full training data and fixed the training epochs to 50. To enable statistical significance testing, we repeated all experiments 10 times with different random seeds, and reported the averaged scores and their standard deviation in all tables.

D. Results

As shown in Table II, we compared our proposed method to state-of-the-art studies in the *short-term glucose prediction* field (Section II-A) and the *postprandial glucose prediction* field (Section II-B). Moreover, the main body of our comparison was our own implementation of several baselines using traditional ML algorithms.

a) Comparing to Short-Term Glucose Prediction Studies: As shown in the first section of Table II, the studies [4]–[6] were all conducted using the same OhioT1DM dataset while evaluated using RMSE between the actual and predicted glucose value without including the other metrics used in our study as the prediction of abnormal-glycemia events was out of their scope. For RMSE, our study was slightly different to these studies as our regression target was the max and min glucose in the next 30 minutes instead of the exact glucose at 30 minutes ahead. However, our methodology achieved a comparable RMSE to these studies when predicting hyperglycemia and an RMSE significantly better than these studies when predicting hypoglycemia, suggesting a performance no lower than the methodologies proposed in these studies if applied to our task.

b) Comparing to Postprandial Glucose Prediction Studies: Among the postprandial glucose prediction studies in our performance comparison in the second section of Table II, [17] followed the fixed prediction horizon scheme with a 30-minute prediction horizon, while [12], [13] adopted the meal-time prediction scheme. These studies were conducted on datasets unavailable to the community due to restrictions of the clinical trials, which made our comparison less straightforward. However, though a lower DT was presented, our study achieved the state-of-the-art MCC, indicating an promising overall performance of our prediction.

c) Comparing to Our Implemented Baselines: To enhance the comparative assessment, we replicated the methodology proposed in [17] and implemented several classical

TABLE III

ABLATION EXPERIMENTS IN THE FORMAT OF MEAN(STD) OVER 10 REPEATED EXPERIMENTS WITH DIFFERENT RANDOM SEEDS.

	Hyperglycemia						Hypoglycemia					
	RMSE↓	SE↑	SP↑	FA↓	MCC↑	DT↑	RMSE↓	SE↑	SP↑	FA↓	MCC↑	DT↑
Individual Training	18.64 (0.14)	0.64 (0.01)	0.96 (0.01)	0.29 (0.02)	0.62 (0.01)	14.22 (0.28)	13.50 (0.08)	0.40 (0.05)	1.00 (0.00)	0.47 (0.05)	0.45 (0.03)	10.13 (0.79)
Binary Supervision	-	0.80 (0.02)	0.80 (0.01)	0.60 (0.01)	0.47 (0.01)	18.11 (0.54)	-	0.70 (0.06)	0.90 (0.05)	0.90 (0.04)	0.24 (0.06)	15.7 (0.80)
Short-Term Training	19.67 (0.21)	0.51 (0.02)	0.97 (0.01)	0.25 (0.02)	0.57 (0.01)	11.71 (0.38)	12.88 (0.15)	0.30 (0.04)	1.00 (0.01)	0.52 (0.09)	0.37 (0.03)	7.13 (0.98)
Ours	18.23 (0.35)	0.66 (0.04)	0.95 (0.01)	0.33 (0.03)	0.61 (0.02)	15.01 (0.98)	13.25 (0.17)	0.48 (0.05)	0.99 (0.01)	0.50 (0.05)	0.48 (0.02)	10.97 (0.89)

ML algorithms as shown in the third section of Table II. All results here were based on the same dataset and experimental conditions. To be specific, we first implemented a dummy model, which simply treated the most recent known glucose value as the regression output $\hat{g}_{\text{hyper}}$ and $\hat{g}_{\text{hypo}}$. As the dummy model had no ability to learn anything, no 10-fold experiment was conducted. This model was used to set up the theoretical lower bound of performance and to show that other methods had the learning ability to an extent. On this basis, we implemented several traditional regression algorithms, namely random forest (RF), AdaBoost, and multi-layer perceptron (MLP) using the same 1-hour past CGM sequence as input, and further replicated the methodology proposed by [17] which was an RF model using explicit features calculated out of the past CGM. In comparison with these baseline strategies, our methodology achieved the highest MCC and lowest RMSE, and comparable performance on other metrics.

d) Discussion: In the OhioT1DM dataset, postprandial hyperglycemia happened around 5 times more often than hypoglycemia which coincides with hypoglycemia usually causing more severe and acute symptoms than hyperglycemia, leading to T1DM patients being more cautious in preventing hypoglycemia. The more significant class imbalance problem in hypoglycemia data caused an overall worse performance in hypoglycemia prediction task (0.13 in MCC). However, this performance gap implied a possibility of having hypoglycemia prediction performance comparable to hyperglycemia prediction by collecting more data as the two tasks only differed in the threshold and time range. Besides, we observed the positive correlation between SE and FA and the trade-off between FA and DT. These observations were understandable because a higher SE indicated a higher tendency of the model to predict the presence of an event, which would also increase the FA. Since the DT was determined by the earliest detection of the event, the more often predicted events would also achieve a better DT. We achieved an average DT of 15 minutes for hyperglycemia and 11 minutes for hypoglycemia. Due to that rapid-acting insulin can be effective in 10-15 minutes [28], and the 20-minute expected recovery time of hypoglycemia [29], we consider our achieved DT to be a promising performance in practical application.

E. Ablation Studies

To further evaluate the validity of our method, we conducted three ablation studies targeting on different arguments in our method (Table III).

a) Individual Training: This ablation experiment aimed at investigating the validity of our joint model for hyperglycemia and hypoglycemia against treating the two tasks separately. Thus, instead of training the model alternatively using data from the two tasks, we trained and evaluated two models individually. In comparing our main experiment to the first row of Table III, our unified model achieved results close to the individual training scheme in all metrics. This observation suggested the capability of the shared LSTM backbone to properly encode the temporal feature of the input glucose history which was then used for prediction on the two tasks. By unifying the two tasks into one model, we minimised the need for computational resources both in inference time and memory requirement.

b) Binary Supervision: Different from previous studies [12], [13], [17], our method approached the classification output in the way of translating the predicted max or min glucose in the prediction horizon to binary result using the defined thresholds. To investigate its effectiveness, we conducted the ablation study of applying binary supervision. More specifically, we changed the output module of our model to a binary classification layer, and the training was supervised by cross-entropy loss with class weights estimated via the class distribution in training data. As shown in the second row of Table III, though the DT was increased, this training strategy led to almost double FA and caused a significant drop of overall predictive performance ($\sim$12% MCC), indicating the numerical supervision being more effective.

c) Short-Term Training: In this study, we treated postprandial glucose prediction in separate to general glucose prediction due to the distribution shift from postprandial glucose to overall glucose as shown in Figure 1 and Table I. To further investigate the effectiveness of this decision, we conducted the ablation experiment of training the model using examples extracted from full CGM data which was consistent to the setting of short-term glucose prediction, yet the testing data were kept unchanged as only postprandial. As can be seen from the third row of Table III, this change not only did not help on increasing the overall performance, but rather caused a deterioration in both the MCC and DT. Additionally, the full training data were approximately of 10 times larger amount than postprandial training data, meaning that invoking a large amount of less relevant training data had only misled the learning.

V. Limitation and Future Work

One limitation of this study is that we did not include meal-related factors such as carbohydrate amount and insulin dosing into our model. Discussing the trade-off between 1) simpler input leads to better applicability, and 2) more complex input provides more potential for better performance, is a good topic for future research. Moreover, future research in direction of a long-term learning model that continuously improves itself using newly collected data from daily life is of great value.

VI. Conclusion

This study identifies the distribution shift of glucose data in the postprandial scenario, and focuses on the problem of postprandial hyperglycemia and hypoglycemia prediction. For the first time we formulate the problem of joint prediction of postprandial hyperglycemia and hypoglycemia. On this basis, we propose a unified ML model that handles the two tasks together and can learn from the glucose pattern only without the need for additional inputs such as meal nutrition in detail. Our experiments on the OhioT1DM dataset achieve state-of-the-art capability to predict postprandial hyperglycemia and hypoglycemia in comparison with existing studies.

Acknowledgements

This research has been delivered in partnership with and funded by Our Health in Our Hands (OHIOH), a strategic initiative of the ANU, which aims to transform health care by developing new personalised health technologies and solutions in collaboration with patients, clinicians, and healthcare providers. We gratefully acknowledge the funding from the ANU School of Computing for the first author's PhD studies.

References

[1] T. Danne, R. Nimri, T. Battelino, R. M. Bergenstal, K. L. Close, J. H. DeVries, S. Garg, L. Heinemann, I. Hirsch, S. A. Amiel *et al.*, "International consensus on use of continuous glucose monitoring," *Diabetes Care*, vol. 40, no. 12, pp. 1631–1640, 2017.

[2] S. Oviedo, J. Vehí, R. Calm, and J. Armengol, "A review of personalized blood glucose prediction strategies for t1dm patients," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 33, no. 6, p. e2833, 2017.

[3] V. Felizardo, N. M. Garcia, N. Pombo, and I. Megdiche, "Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction–a systematic literature review," *Artificial Intelligence in Medicine*, vol. 118, p. 102120, 2021.

[4] Y. Deng, L. Lu, L. Aponte, A. M. Angelidi, V. Novak, G. E. Karniadakis, and C. S. Mantzoros, "Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients," *NPJ Digital Medicine*, vol. 4, no. 1, pp. 1–13, 2021.

[5] T. Yang, X. Yu, N. Ma, R. Wu, and H. Li, "An autonomous channel deep learning framework for blood glucose prediction," *Applied Soft Computing*, vol. 120, p. 108636, 2022.

[6] R. Cui, C. Hettiarachchi, C. J. Nolan, E. Daskalaki, and H. Suominen, "Personalised short-term glucose prediction via recurrent self-attention network," in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2021, pp. 154–159.

[7] T. Zhu, K. Li, P. Herrero, J. Chen, and P. Georgiou, "A deep learning algorithm for personalized blood glucose prediction." in *The 3rd International Workshop on Knowledge Discovery in Healthcare Data@IJCAI-ECAI*, 2018, pp. 64–78.

[8] T. Yang, R. Wu, R. Tao, S. Wen, N. Ma, Y. Zhao, X. Yu, and H. Li, "Multi-scale long short-term memory network with multi-lag structure for blood glucose prediction." in *5th International Workshop on Knowledge Discovery in Healthcare Data*, 2020, pp. 136–140.

[9] J. Martinsson, A. Schliep, B. Eliasson, and O. Mogren, "Blood glucose prediction with variance estimation using recurrent neural networks," *Journal of Healthcare Informatics Research*, vol. 4, no. 1, pp. 1–18, 2020.

[10] C. Marling and R. Bunescu, "The OhioT1DM dataset for blood glucose level prediction: Update 2020," in *CEUR Workshop Proceedings*, vol. 2675. NIH Public Access, 2020, p. 71.

[11] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[12] S. Oviedo, I. Contreras, C. Quirós, M. Giménez, I. Conget, and J. Vehi, "Risk-based postprandial hypoglycemia forecasting using supervised learning," *International Journal of Medical Informatics*, vol. 126, pp. 1–8, 2019.

[13] J. Vehí, I. Contreras, S. Oviedo, L. Biagi, and A. Bertachi, "Prediction and prevention of hypoglycaemic events in type-1 diabetic patients using machine learning," *Health Informatics Journal*, vol. 26, no. 1, pp. 703–718, 2020.

[14] D. Adelberger, F. Reiterer, P. Schrangl, C. Ringemann, T. Huschto, and L. del Re, "Prediction of postprandial glucose excursions in type 1 diabetes using control-oriented process models," *IFAC-PapersOnLine*, vol. 54, no. 15, pp. 466–471, 2021.

[15] E. A. Pustozerov, A. S. Tkachuk, E. A. Vasukova, A. D. Anopova, M. A. Kokina, I. V. Gorelova, T. M. Pervunina, E. N. Grineva, and P. V. Popova, "Machine learning approach for postprandial blood glucose prediction in gestational diabetes mellitus," *IEEE Access*, vol. 8, pp. 219 308–219 321, 2020.

[16] R. A. Karim, I. Vassányi, and I. Kósa, "After-meal blood glucose level prediction using an absorption model for neural network training," *Computers in Biology and Medicine*, vol. 125, p. 103956, 2020.

[17] W. Seo, Y.-B. Lee, S. Lee, S.-M. Jin, and S.-M. Park, "A machine-learning approach to predict postprandial hypoglycemia," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1–13, 2019.

[18] S. Daenen, A. Sola-Gazagnes, J. M'Bemba, C. Dorange-Breillard, F. Defer, F. Elgrably, E. Larger, and G. Slama, "Peak-time determination of post-meal glucose excursions in insulin-treated diabetic patients," *Diabetes & Metabolism*, vol. 36, no. 2, pp. 165–169, 2010.

[19] Y. Altuntaş, "Postprandial reactive hypoglycemia," *Şişli Etfal Hastanesi tıp Bülteni*, vol. 53, no. 3, p. 215, 2019.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] S. Cichosz, "Data sharing of continuous glucose monitoring data: The need for a new paradigm?" *Journal of Diabetes Science and Technology*, vol. 15, no. 4, p. 961, 2021.

[22] C. Hettiarachchi, E. Daskalaki, J. Desborough, C. J. Nolan, D. O'Neal, H. Suominen *et al.*, "Integrating multiple inputs into an artificial pancreas system: Narrative literature review," *JMIR Diabetes*, vol. 7, no. 1, p. e28861, 2022.

[23] C. Hettiarachchi, N. Malagutti, C. Nolan, E. Daskalaki, and H. Suominen, "A reinforcement learning based system for blood glucose control without carbohydrate estimation in type 1 diabetes: In silico validation," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 950–956.

[24] R. M. Bergenstal, A. J. Ahmann, T. Bailey, R. W. Beck, J. Bissen, B. Buckingham, L. Deeb, R. H. Dolin, S. K. Garg, R. Goland *et al.*, "Recommendations for standardizing glucose reporting and analysis to optimize clinical decision making in diabetes: the ambulatory glucose profile," *Diabetes Technology Society*, 2013.

[25] N. Brew-Sam, M. Chhabra, A. Parkinson, K. Hannan, E. Brown, L. Pedley, K. Brown, K. Wright, E. Pedley, C. J. Nolan *et al.*, "Experiences of young people and their caregivers of using technology to manage type 1 diabetes mellitus: Systematic literature review and narrative synthesis," *JMIR Diabetes*, vol. 6, no. 1, p. e20973, 2021.

[26] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *ICML*, 2010.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[28] S. K. Garg, S. L. Ellis, and H. Ulrich, "Insulin glulisine: a new rapid-acting insulin analogue for the treatment of diabetes," *Expert Opinion on Pharmacotherapy*, vol. 6, no. 4, pp. 643–651, 2005.

[29] A. Nakhleh and N. Shehadeh, "Hypoglycemia in diabetes: An update on pathophysiology, treatment, and prevention," *World Journal of Diabetes*, vol. 12, no. 12, p. 2036, 2021.