Prediction of Maximum and Minimum Postprandial Blood Glucose Levels in People with Diabetes

Kotaro Nagayama*, Shota Kato*, Kana Eguchi*, Masahide Hamaguchi†,
Hiroyuki Tominaga†, Youji Hamaguchi‡, Michiaki Fukui†, and Manabu Kano*
* Kyoto University, Japan
E-mail: nagayama.kotaro.63c@st.kyoto-u.ac.jp, shota@human.sys.i.kyoto-u.ac.jp,
kana.eguchi@ieee.org, manabu@human.sys.i.kyoto-u.ac.jp.
† Kyoto Prefectural University of Medicine, Japan
E-mail: mhama@koto.kpu-m.ac.jp, htommy@koto.kpu-m.ac.jp, michiaki@koto.kpu-m.ac.jp.
‡ S.I.P, Japan
E-mail: hamaguchi@sipartners.jp

***Abstract*—Diabetes mellitus (DM) is a chronic disease that elevates blood glucose levels and increases cardiovascular risks. Clinical care focuses on limiting postprandial glucose excursions within the first three hours after a meal while avoiding hypoglycemia, but predicting these fluctuations remains challenging. We propose a deep neural network based on the Temporal Fusion Transformer (TFT) architecture to predict postprandial maximum and minimum glucose levels. The model integrates four types of information: continuous blood glucose measurements, meal details, insulin doses, and individual attributes. We collected data from 56 individuals with DM and divided them into training, validation, and test sets for each individual. The proposed model achieved mean absolute errors of 29.7 mg/dL for postprandial maxima and 20.2 mg/dL for minima, with 61.3% and 78.6% of predictions within ±30 mg/dL, respectively. Although further improvements are necessary for real-world application, this work introduces a novel framework for directly predicting postprandial glucose extremes and provides detailed analyses by DM types to guide future clinical interventions.**

I. Introduction

Diabetes mellitus (DM) is a disease characterized by chronic hyperglycemia resulting from insufficient insulin action [1]. Large fluctuations in blood glucose levels are known to promote the progression of arteriosclerosis and increase the risk of cardiovascular disease [2], [3]; thus, controlling these fluctuations is essential.

Insulin therapy is a key treatment for DM. It adjusts blood glucose levels by administering insulin externally. Optimal insulin doses vary by person and by various factors such as meals and exercise. Thus, people with DM are required to determine the dose before mealtimes that is appropriate for minimizing blood glucose fluctuations after the meal. A recommended strategy is carbohydrate counting, which estimates an insulin dose based on the total carbohydrate content in an upcoming meal. However, this method requires practice and skill. Proficiency with carbohydrate counting varies among individuals and may require support or training [4]. In addition, inaccurate dosing carries the risk of hyperglycemia or hypoglycemia, both of which are serious health concerns. This context has necessitated a systematic and user-friendly method for determining the appropriate dosage.

Recent studies have applied machine learning techniques to automate insulin dose estimation [5]–[8]. Gupta et al. [7] developed a model to predict future insulin doses from historical blood glucose and insulin dosing data using long short-term memory (LSTM) [9]. Jisha et al. [8] used the OhioT1DM dataset [10] to build a predictive model based on features such as meal information, exercise levels, and insulin doses, achieving higher accuracy than Gupta et al. [7]. However, existing approaches are designed to reproduce historical clinician decisions, which may not always reflect the optimal dosing for each individual. Although reproducing prior data may reduce clinical workload, it does not ensure that the recommended doses are systematically optimized for individuals. In contrast, our study aims to directly predict postprandial glucose extremes and provide a framework for systematically deriving dosing strategies that minimize glucose variability.

The goal of this research is to develop a method for computing insulin doses that minimize blood glucose fluctuations after a meal. As its initial step, this paper focuses on accurately predicting the maximum and minimum postprandial blood glucose values within a three-hour period. We propose a method that directly predicts these postprandial extremes of glucose, and the methodology appears in Section II. This paper discusses the effectiveness of the proposed approach and related challenges.

From a clinical perspective, an accurate prediction of minimum postprandial glucose values is more important than that of maximum values, because hypoglycemia below 70 mg/dL can have adverse effects on the body. Due to the inherent errors in many blood glucose measurement devices, the general clinical goal is that 90% of predicted values should fall within ±30 mg/dL of the actual value. These two thresholds motivate the direct prediction of the postprandial blood glucose extremes themselves.

979-8-3315-7206-8/25/$31.00

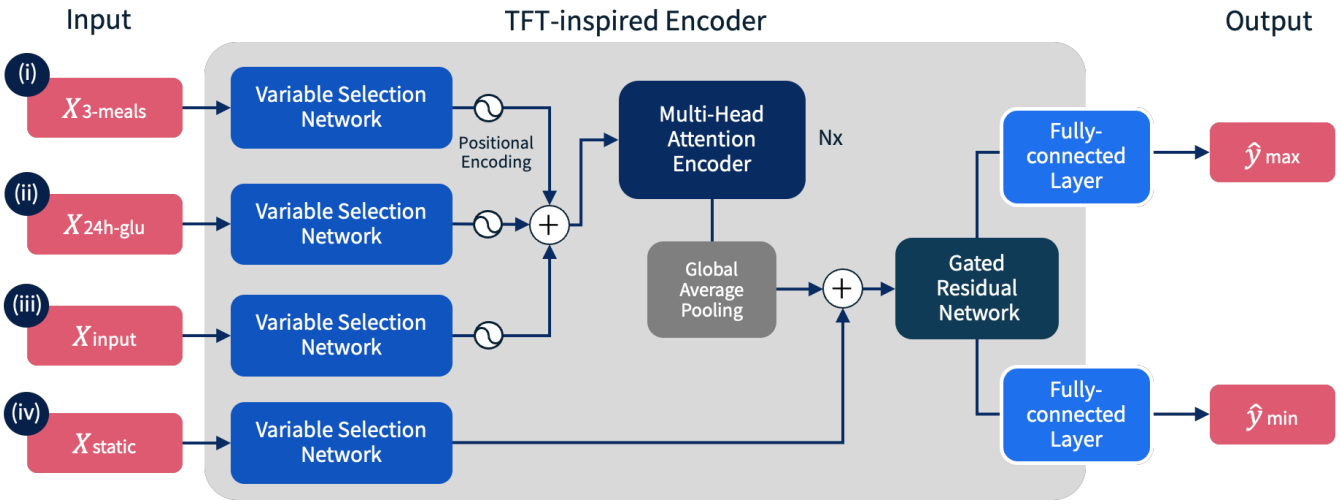

Fig. 1. Overview of the proposed model. The model takes as inputs (i) data on the past three meals, (ii) blood glucose levels in the preceding 24 hours, (iii) current meal and blood glucose data, and (iv) individual attributes. Nx denotes the number of stacked multi-head attention blocks, and TFT is the temporal fusion transformer. The output is the prediction of the postprandial glucose maximum and minimum values.

II. Materials And Methods

This study analyzes blood glucose data, insulin doses, meal information, and individual attributes collected from people with DM. We build and test a model to predict maximum and minimum postprandial glucose within three hours after a meal.

The protocol of the study was approved by the Medical Ethics Review Committee, Kyoto Prefectural University of Medicine (ERB-C-2553).

A. Dataset

This study uses records from 56 individuals who received care at Kyoto Prefectural University of Medicine Hospital and wore continuous glucose monitors for an average of 100 days; 31 individuals were women, 25 were men, and the median age was 56 years with an interquartile range of 48 to 66 years. In terms of DM type, the cohort comprised 42 individuals with type-1 DM (T1DM), 11 with type-2 DM (T2DM), and three with steroid-induced DM (SIDM).

The data consist of the following four features:

- **individual attributes:** the sex, age, and DM type
- **Blood glucose:** continuously measured values retrieved at intervals of 5 to 15 minutes
- **Meal information:** the amount of carbohydrates and the meal type (breakfast, lunch, or dinner)
- **Insulin dose:** bolus insulin administered immediately before a meal

B. Data Preprocessing

Each individual's blood glucose data are resampled at five-minute intervals. Missing values are linearly interpolated. Then, for each meal, the interval from mealtime start to three hours later is segmented as one set. Those sets are arranged in chronological order for each individual.

C. Proposed Model

Fig. 1 illustrates the proposed model. We employ an encoder inspired by the temporal fusion transformer (TFT) [11] to extract features from historical data. TFT is a derivative of the Transformer [12] architecture specifically tailored for time series forecasting. In addition to the standard multi-head self-attention mechanism, TFT incorporates a variable selection network that learns to weigh each input feature according to its predictive importance, and it employs gated residual connections to regulate the flow of information through the network. These enhancements allow TFT to handle heterogeneous data streams including known future covariates and static individual attributes. For these reasons, TFT is especially well suited to the task of directly predicting postprandial glucose extremes.

The inputs comprise four groups of features:

(i) $X_{\text{3-meals}} \in \mathbb{R}^{T_\text{p} \times F_\text{p}}$: This represents **data on the past three meals**. Information on each meal includes carbohydrate intake, insulin dose, and blood glucose fluctuations for three hours after the meal. This set of features captures the relationship among meals, insulin doses, and postprandial glucose. Here, T_p is the total number of time steps across the past three meals and F_p is the number of features per segment.

(ii) $X_{\text{24h-glu}} \in \mathbb{R}^{T_\text{g} \times F_\text{g}}$: This represents **blood glucose levels in the preceding 24 hours**. This set of features captures the daily patterns of blood glucose. Here, T_g is the number of five-minute time steps in the preceding 24 hours and F_g is the number of glucose-related features per step.

(iii) $X_{\text{input}} \in \mathbb{R}^{F_\text{f}}$: This represents **current meal and blood glucose data**. This set of features includes the planned carbohydrate intake, the expected insulin dose, and the current blood glucose level, providing information about upcoming meals and insulin doses. The number of these features is denoted by F_f.

(iv) $X_{\text{static}} \in \mathbb{R}^{F_s}$: This represents **individual attributes**. This set of features consists of the sex, age, and DM type. Here, F_s is the number of static individual attributes.

We set $T_p = 108$, $F_p = 4$, $T_g = 288$, $F_g = 1$, $F_f = 4$, and $F_s = 3$ in this study. The model accepts four distinct input sequences and directly predicts the postprandial glucose maximum $\hat{y}_{\max}$ and minimum $\hat{y}_{\min}$ by two separate multi-layer perceptron (MLP) regression heads. Each input is processed in the following four steps:

1) **Feature weighting**: The variable selection networks (VSN) first project each feature to a common embedding dimension d_{model} by a linear layer. It then uses a gated residual network (GRN) to compute a normalized weight for each feature, applies these weights to the embeddings, and sums across features. Through this procedure, the model obtains an embedding vector $\tilde{X}_t \in \mathbb{R}^{d_{\text{model}}}$ for each time step t. This mechanism allows the model to suppress noisy or irrelevant signals dynamically and focus on the most predictive inputs at each time step.
   Each GRN block performs a non-linear transformation with residual gating to preserve information flow. Internally, it consists of the following four components:
   - a linear projection to a hidden dimension, followed by exponential linear unit (ELU) activation and dropout
   - a second linear projection back to the original embedding dimension
   - a gated linear unit (GLU) that learns to combine the transformed signal with the original input
   - a final layer normalization to stabilize training

   This mechanism enables both complex feature learning and unimpeded gradient flow and thereby enhances model stability.
2) **Time-series fusion and encoding**: The model incorporates a dedicated positional encoding into each embedding vector $\tilde{X}$ before the three sequences are concatenated.
   Each positional encoding $P \in \mathbb{R}^{I \times d_{\text{model}}}$ has I rows for time-step indices $i \in \{0, \ldots, I-1\}$ and d_{model} columns for embedding indices $j \in \{0, \ldots, d_{\text{model}} - 1\}$. Each element is

$$P_{i,j} = \begin{cases} \sin\big(i/10000^{j/d_{\text{model}}}\big) & \text{if } j \bmod 2 = 0, \\ \cos\big(i/10000^{(j-1)/d_{\text{model}}}\big) & \text{if } j \bmod 2 = 1. \end{cases} \tag{1}$$

   The positional encodings assigned to the three sequences are denoted by

$$P_{\text{3-meals}} \in \mathbb{R}^{T_p \times d_{\text{model}}}, \tag{2}$$
$$P_{\text{24h-glu}} \in \mathbb{R}^{T_g \times d_{\text{model}}}, \tag{3}$$
$$P_{\text{input}} \in \mathbb{R}^{d_{\text{model}}}. \tag{4}$$

   The encoded sequences are then formed as

$$\hat{X}_{\text{3-meals}} = \widetilde{X}_{\text{3-meals}} + P_{\text{3-meals}}, \tag{5}$$
$$\hat{X}_{\text{24h-glu}} = \widetilde{X}_{\text{24h-glu}} + P_{\text{24h-glu}}, \tag{6}$$
$$\hat{X}_{\text{input}} = \widetilde{X}_{\text{input}} + P_{\text{input}}. \tag{7}$$

   These three encoded sequences are concatenated as

$$H = \text{concat}(\hat{X}_{\text{3-meals}}, \hat{X}_{\text{24h-glu}}, \hat{X}_{\text{input}}) \in \mathbb{R}^{(T_p+T_g+1) \times d_{\text{model}}}. \tag{8}$$

   The resulting H is passed through L layers of gated multi-head attention to capture temporal dependencies.
3) **Context vector generation**: The output of the final attention layer is average-pooled over time to yield $H_{\text{seq}} \in \mathbb{R}^{d_{\text{model}}}$. The static input is separately embedded to $H_s \in \mathbb{R}^{d_{\text{model}}}$, and these are combined by a GRN to obtain the final context vector H_{ctx}:

$$H_{\text{ctx}} = \text{GRN}\big(\alpha\, H_{\text{seq}} + (1-\alpha)\, H_s\big). \tag{9}$$

   Here, $\alpha \in [0, 1]$ is a parameter that balances the contribution of dynamic sequence features and static features.
4) **Dual regression heads**: The context vector H_{ctx} is fed to two MLPs to predict the maximum and minimum postprandial glucose:

$$\hat{y}_{\max} = \text{MLP}_{\max}(H_{\text{ctx}}), \quad \hat{y}_{\min} = \text{MLP}_{\min}(H_{\text{ctx}}). \tag{10}$$

   Each of the two multi-layer perceptron heads $\text{MLP}_{\max}$ and $\text{MLP}_{\min}$ consists of three fully-connected layers.

D. Experimental Settings

The data are split into training, validation, and test sets for each individual in chronological order, which comprise 1169, 328, and 351 samples, respectively. There is no overlap among the samples in these three sets. We set d_{model} to 256, L to six, attention heads to eight, the learning rate to 3×10^{-5} and choose AdamW [13] as the optimizer. We use Optuna [14] to optimize the dropout rate and α. The model is trained over a total of 100 epochs, with the validation loss being calculated at each epoch. The model with the lowest validation loss is then adopted for the test dataset. We employ the mean squared error (MSE) as the loss function:

$$\frac{1}{N} \sum_{n=1}^{N} \{(\hat{y}_{\max}^{(n)} - y_{\max}^{(n)})^2 + (\hat{y}_{\min}^{(n)} - y_{\min}^{(n)})^2\}. \tag{11}$$

Here, $\hat{y}_{\max}^{(n)}$, $y_{\max}^{(n)}$, $\hat{y}_{\min}^{(n)}$, and $y_{\min}^{(n)}$ represent the n-th maximum and minimum values of the predicted and measured values, respectively, and N is the number of training samples.

In this experiment, we address two tasks: the maximum-glucose task and the minimum-glucose task. The former assesses the model's prediction of the maximum glucose value within three hours after a meal, whereas the latter assesses the model's prediction of the minimum value within the same window. Furthermore, in order to compare the prediction performance among different types of DM, we perform a chi-square test of the relationship between prediction success and type of DM.

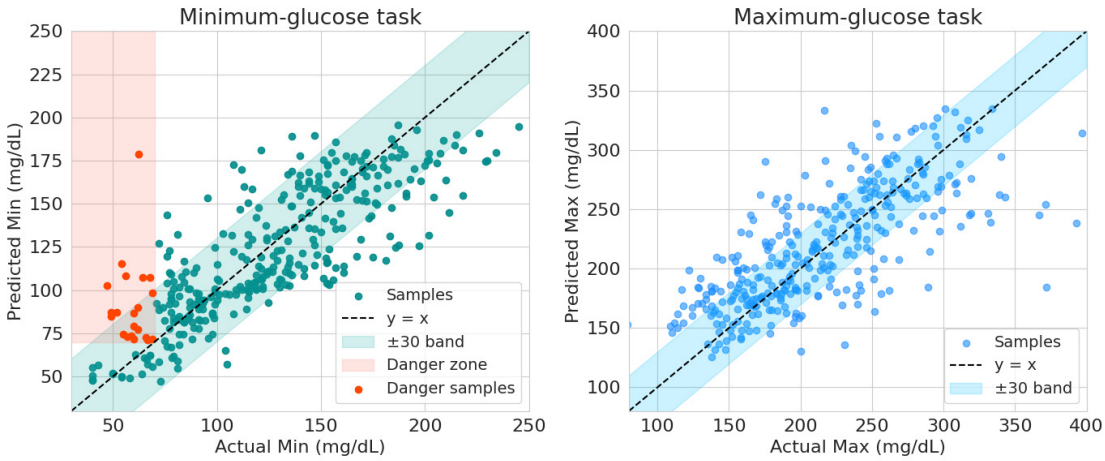

Fig. 2. Scatter plot of measured and predicted minimum and maximum values. The x-axis represents measured values, the y-axis represents predicted values, and the shaded areas indicate ±30 mg/dL from measured values. On the left-hand side (minimum-glucose task), the red area indicates an area where the risk of hypoglycemia is not predicted.

E. Evaluation Metrics

Prediction performance is quantified using two metrics. One is the mean absolute error (MAE) between the predicted and observed glucose extremes:

$$\mathrm{MAE} = \frac{1}{N}\sum_{n=1}^{N}\left|\hat{y}_{\mathrm{ext}}^{(n)} - y_{\mathrm{ext}}^{(n)}\right| \quad \mathrm{ext} \in \{\max, \min\}. \tag{12}$$

The other is the proportion of predictions falling within ±30 mg/dL of the measured values ($\mathrm{Coverage}_{30}$) :

$$\mathrm{Coverage}_{30} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{1}\left(\left|\hat{y}_{\mathrm{ext}}^{(n)} - y_{\mathrm{ext}}^{(n)}\right| \le 30\right) \quad \mathrm{ext} \in \{\max, \min\}, \tag{13}$$

where $\mathbf{1}(\cdot)$ is the indicator function. The threshold value ±30 mg/dL represents the maximum error that clinicians consider tolerable in practice. These metrics are applied separately to the maximum-glucose task and the minimum-glucose task.

III. Results and Discussion

A. Overall Prediction Results

Fig. 2 compares the predicted and measured extremes. The shaded bands mark the region in which the absolute prediction error is not greater than 30 mg/dL. In the maximum-glucose task, 61.3% of the points lay inside the band, and in the minimum-glucose task, 78.6% of the points lay inside the band.

Fig. 3 presents boxplots of the absolute prediction error across all test samples. The left boxplot refers to the minimum-glucose task and the right boxplot to the maximum-glucose task. The median error for the minimum values was 15.5 mg/dL, whereas the median error for the maximum values was 23.1 mg/dL. The maximum-glucose boxplot had a wider interquartile range and a larger upper whisker than the minimum-glucose boxplot.

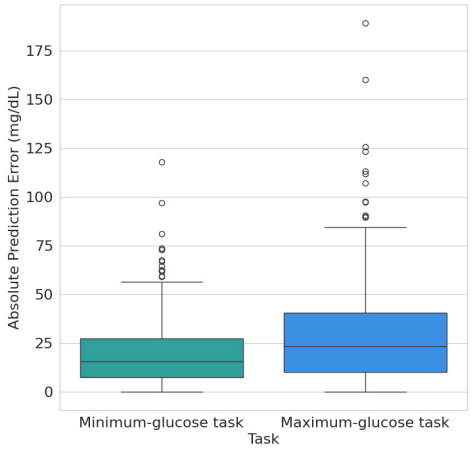

Fig. 3. Distribution of absolute prediction errors for postprandial glucose maximum value and minimum value.

TABLE I
Prediction Performance for Postprandial Glucose Extremes

	MAE (mg/dL)	$\mathrm{Coverage}_{30}$ (%)
Minimum-glucose task	20.2	78.6
Maximum-glucose task	29.7	61.3

Table I lists the MAE and the $\mathrm{Coverage}_{30}$. The MAE equaled 20.2 mg/dL for the minimum values and 29.7 mg/dL for the maximum values. The higher accuracy in the minimum glucose task can be attributed to its narrower range of values compared to the maximum glucose task.

B. Minimum Value Prediction and Hypoglycemia

Overestimating the minimum value may lead to underestimating the risk of hypoglycemia. If the true value is less than 70 mg/dL, overestimating it may mask the risk of hypoglycemia. Therefore, it is necessary to consider the sign

TABLE II
PREPRANDIAL GLUCOSE STATISTICS

Group	Samples	Mean (mg/dL)	Median (mg/dL)
Danger	19	120	106
Normal	332	153	151

of the error as well as evaluate the absolute prediction error when comparing predicted values with actual measurements.

On the left side of Fig. 2, the red area indicates an area where the measured value is below 70 mg/dL and the predicted value is above 70 mg/dL. In the test set, 19 of 351 meal episodes satisfied this criterion, which corresponds to 5.4 %. Further improvements are needed for practical use. Table II compares the preprandial glucose values for the danger and normal groups. The mean of the danger group was 33 mg/dL lower than that of the normal group, and its median was 45 mg/dL lower. This indicates that the model is prone to overestimate the minimum value when the preprandial glucose is in a low range. As a potential solution, additional penalization for such instances could be incorporated into the model to improve predictions in these critical regions.

C. Relationship between Prediction Accuracy and DM Types

This subsection focuses on the DM type as a potential determinant of prediction success. Fig. 4 visualizes the absolute error distributions for T1DM, T2DM, and SIDM in both the minimum-glucose task and the maximum-glucose task. To assess whether these observed differences are statistically meaningful, each meal was classified as *Success* if its absolute error was within 30 mg/dL and as *Fail* otherwise, and a 3×2 chi-square test was performed.

As shown in Table III, in the minimum-glucose task, the rates of *Success* achieved 81%, 56%, and 83% for T1DM, T2DM, and SIDM meals, respectively. In the maximum-glucose task, the corresponding rates were 60%, 58%, and 91%.

In the minimum-glucose task, the chi-square statistic equaled 9.42 with two degrees of freedom and the p-value was 0.009, leading to rejection of the null hypothesis of independence. In the maximum-glucose task, no significant association was detected, since the chi-square statistic was 4.93 and the p-value was 0.085.

We attribute the significant association for minimum values to physiological factors. The lowest post-meal glucose level depends on hepatic glucose release, insulin sensitivity, and other mechanisms. These mechanisms vary across DM types and are not captured by the current feature set and therefore the model cannot adjust for them.

We attribute the lower accuracy in T2DM to two factors. First, the wide range of insulin sensitivity in T2DM can result in different glucose trajectories for the same insulin dose and carbohydrate. Second, the heterogeneity in treatment patterns and physiological responses in people with T2DM presents unique challenges for predictive modeling. T1DM and SIDM have more uniform pharmacodynamic responses, which may contribute to their higher accuracy.

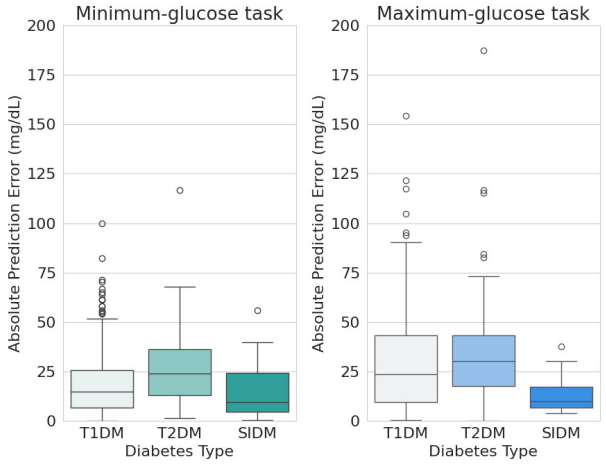

Fig. 4. Absolute prediction error by DM type. The left panel shows the minimum-glucose task and the right panel the maximum-glucose task.

TABLE III
PREDICTION SUCCESS RATE BY DM TYPE

DM type	Number of meals	Success rate (%)	
		Minimum	Maximum
T1DM	136	81	60
T2DM	157	65	58
SIDM	58	83	91

D. Limitation

This study has several limitations. One limitation is the tendency to underestimate dangerously low blood glucose values as discussed in III-B. It can mask potential hypoglycemia and lead to missed alerts. Another limitation is the data used for training. Meal and insulin dosage records rely on individual self-reporting, which can introduce inaccuracies. Specifically, some meals were recorded later than the actual time of intake, and some records were missing. The blood glucose measurement device can also produce measurement errors. The absence of physical activity data may affect predictions because exercise has a significant impact on postprandial glucose levels.

Future work will address these issues by refining the loss function to penalize errors more strictly in the hypoglycemic range, and by exploring more comprehensive data collection methods to improve robustness.

IV. CONCLUSION

In this study, we developed a deep learning model that employs a temporal fusion transformer encoder to predict the maximum and minimum blood glucose values during the three hours following a meal for people with DM. The model uses continuous glucose measurements, meal information, insulin doses, and individual attributes collected from individuals with DM as input features. This study made a novel contribution

by introducing a deep learning-based approach for direct prediction of postprandial maximum and minimum blood glucose levels. The study also provided insights into how prediction errors vary by DM types, and evaluated the model from a clinical perspective by applying safety-relevant metrics to assess its potential for real-world use.

A key challenge remains in addressing cases where actual measurements indicate a risk of hypoglycemia but the proposed method does not predict such a risk. Moreover, future work will develop a system to recommend insulin doses that minimize postprandial glucose fluctuations by leveraging this model.

Acknowledgement

The authors appreciate Kyoto Medical Innovation Research Program, Circulating Research Grant from the Certified NPO Japan IDDM Network, Feasibility Study Grant [Startup Challenge Type] from the Japan Science and Technology Agency (JST), Medical Innovation Research Grant, R&D Support Program for Growth-Oriented SMEs, and Tanuma Greenhouse Foundation Research Grant.

References

[1] L. Ministry of Health and Welfare, "Guide to insulin dosage and carbohydrate counting," Ministry of Health, Labour and Welfare, Tech. Rep., 2021, Technical guide.

[2] S. Wakasugi, T. Mita, N. Katakami, *et al.*, "Associations between continuous glucose monitoring-derived metrics and arterial stiffness in japanese patients with type 2 diabetes," *Cardiovascular Diabetology*, vol. 20, p. 15, 2021.

[3] A. Kitae, T. Kimura, C. Munekawa, *et al.*, "Development of application to automatically calculate mean amplitude of glycaemic excursions using intermittently scanned continuous glucose monitoring data," *Diabetes, Obesity and Metabolism*, vol. 23, no. 9, pp. 2155–2160, 2021.

[4] L. T. Meade and W. E. Rushton, "Accuracy of carbohydrate counting in adults," *Clinical Diabetes*, vol. 34, no. 3, pp. 142–147, 2016.

[5] S. A. Neethika, K. V. V. S. L. Varshitha, L. N. Jahnavi, N. Mounika, and N. M. Sowjanya, "An expert system for insulin dosage prediction," *International Journal of Advanced Research in Science, Communication and Technology*, vol. 4, no. 5, pp. 215–222, 2024.

[6] V. Anirudh, G. Siri, P. Shashank, T. R. Vardhan, and V. B. Rao, "Prediction of diabetes and insulin dosage," *International Journal of Advances in Engineering and Management*, vol. 5, no. 4, pp. 877–883, 2023.

[7] K. Gupta and N. Jiwani, "Prediction of insulin level of diabetes patient using machine learning approaches," *International Journal of Creative Research Thoughts*, vol. 10, no. 8, pp. C434–C441, 2022.

[8] G. Jisha, N. T. Bhuvan, and R. Jerrard, "Time series prediction of personalized insulin dosage for type 2 diabetics," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 2, pp. 1080–1087, 2023.

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

[10] C. Marling and R. Bunescu, "The ohiot1dm dataset for blood glucose level prediction: Update 2020," in *CEUR Workshop Proceedings*, 2020.

[11] B. Lim, S. O. Arik, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.

[12] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[13] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[14] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2019, pp. 2623–2631.