

Exploring the potential of deep learning models integrating transformer and LSTM in predicting blood glucose levels for T1D patients

Xin Xiong^{1,*} , XinLiang Yang^{1,*} , Yunying Cai^{2,*} , Yuxin Xue¹ , JianFeng He¹  and Heng Su²

Abstract

Objective: Diabetes mellitus is a chronic condition that requires constant blood glucose monitoring to prevent serious health risks. Accurate blood glucose prediction is essential for managing glucose fluctuations and reducing the risk of hypo- and hyperglycemic events. However, existing models often face limitations in prediction horizon and accuracy. This study aims to develop a hybrid deep learning model combining Transformer and Long Short-Term Memory (LSTM) networks to improve prediction accuracy and extend the prediction horizon, using personalized patient information and continuous glucose monitoring data to support better real-time diabetes management.

Methods: In this study, we propose a hybrid deep learning model combining Transformer and LSTM networks to predict blood glucose levels for up to 120 min. The Transformer Encoder captures long-range dependencies, while the LSTM models short-term patterns. To improve feature extraction, we integrate Bidirectional LSTM and Transformer Encoder layers at multiple stages. We also use positional encoding, dropout layers, and a sliding window technique to reduce noise and manage temporal dependencies. Richer features, including meal composition and insulin dosage, are incorporated to enhance prediction accuracy. The model's performance is validated using real-world clinical data and error grid analysis.

Results: On clinical data, the model achieved root mean square error/mean absolute error of 10.157/6.377 (30-min), 10.645/6.417 (60-min), 13.537/7.283 (90-min), and 13.986/6.986 (120-min). On simulated data, the results were 1.793/1.376 (15-min), 2.049/1.311 (30-min), and 3.477/1.668 (60-min). Clark Grid Analysis showed that over 96% of predictions fell within the clinical safety zone up to 120 min, confirming its clinical feasibility.

Conclusion: This study demonstrates that the combined Transformer and LSTM model can effectively predict blood glucose concentration in type 1 diabetes patients with high accuracy and clinical applicability. The model provides a promising solution for personalized blood glucose management, contributing to the advancement of artificial intelligence technology in diabetes care.

Keywords

type 1 diabetes, predictive analytics, deep learning model, short-term blood glucose prediction, Clarke Error Grid Analysis

Received: 24 October 2024; accepted: 3 March 2025

¹Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China

²Department of Endocrinology, The First People's Hospital of Yunnan Province, The Affiliated Hospital of Kunming University of Science and Technology, Kunming, China

*Xin Xiong, XinLiang Yang and Yunying Cai contributed equally to this work.

Corresponding authors:

JianFeng He, Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China.
Email: jfenghe@kust.edu.cn

Heng Su, Department of Endocrinology, The First People's Hospital of Yunnan Province, The Affiliated Hospital of Kunming University of Science and Technology, Kunming 650500, China.
Email: su_hen@hotmail.com



Introduction

Diabetes mellitus (DM) is a complex metabolic disorder influenced by genetic, immune, and environmental factors. It primarily manifests as insulin secretion deficiencies and/or reduced tissue sensitivity to insulin, disrupting glucose, carbohydrate, lipid, and protein metabolism. Among the various forms of diabetes, type 1 diabetes (T1D) is particularly challenging, as it involves the autoimmune destruction of insulin-producing beta cells. This necessitates lifelong insulin therapy to maintain blood glucose concentrations (BGC) within a therapeutic range, preventing the onset of serious complications such as cardiovascular diseases, retinopathy, nephropathy, and neuropathy.

Maintaining optimal BGC is crucial for managing T1D. Hyperglycemia, defined as blood glucose exceeding 180 mg/dL, can result in severe complications. Such as damage to the eyes, kidneys, and nervous system,¹ while hypoglycemia (blood glucose below 70 mg/dL) caused by excessive insulin can lead to fainting, coma, or even death.² To reduce these risks, precise blood glucose prediction is essential. Accurate predictions help determine the right insulin doses, meal plans, and physical activity regimens, supporting both day-to-day management and long-term health outcomes.

A promising tool for blood glucose monitoring is continuous glucose monitoring (CGM), which provides real-time data on blood glucose levels through small sensors, eliminating the need for painful finger pricks.³ Continuous glucose monitoring devices produce high-density time-series data, offering continuous and detailed measurements of blood glucose.

However, despite the advancements in CGM technologies, accurate blood glucose prediction remains a significant challenge. The dynamics of blood glucose are influenced by a variety of factors, including carbohydrate intake, insulin doses, physical activity, stress, infections, sleep patterns, and hormonal fluctuations.⁴ Additionally, T1D patients show considerable individual variation in their glucose responses to insulin, further complicating predictions. These factors make it difficult to model glucose levels reliably and consistently across all patients.⁵

In recent years, deep learning has emerged as a powerful tool for predicting blood glucose levels in diabetic patients. One of its key advantages is the ability to automatically extract relevant features from raw input data, eliminating the need for manual feature engineering. Deep learning models can handle complex and high-dimensional datasets, support multiple inputs and outputs, and effectively capture patterns in long sequences of data.⁶ These capabilities allow deep learning to provide more accurate and reliable predictions, even in the presence of varying factors such as insulin doses, physical activity, and meal timing. This makes deep learning particularly well-suited for real-time blood glucose monitoring and management.

Bertachi et al.⁷ combined artificial neural networks (ANN) with physiological models to predict blood glucose and detect nighttime hypoglycemia. The model used glucose, insulin, carbohydrate, and activity data, trained on a split dataset. Predictions were evaluated using root mean square error (RMSE). In the OhioT1DM dataset, RMSE for 30-min and 60-min predictions were 19.33 mg/dL and 31.72 mg/dL, respectively. The model achieved 85% sensitivity and 92% specificity for nighttime hypoglycemia. Zhu et al.⁵ used an Extended Recurrent Neural Network (RNN) with vanilla RNN layers for efficient blood glucose prediction within 30 min. The model achieved an RMSE of 18.9 ± 2.6 mg/dL on the OhioT1DM dataset, outperforming NNPG, SVR, and ARX models. Similarly, Cappon et al.⁸ developed a personalized BiLSTM model that processes sequences in both directions, enhancing pattern recognition. This model achieved an RMSE of 20.20 mg/dL for 30-min predictions and 34.19 mg/dL for 60-min predictions on the same dataset.

GluNet⁹ is a deep learning model designed for blood glucose prediction using artificially generated continuous monitoring data from T1D patients. It incorporates preprocessing, label transformation, multilayer dilated convolutions with gate activations, and postprocessing for computational efficiency. The model achieved an RMSE of 19.28 ± 2.76 mg/dL for 30-min predictions and 31.83 ± 3.49 mg/dL for 60-min predictions. Similarly, Mehrad Jaloli et al.¹⁰ proposed a CNN-LSTM model that integrates 1D convolutional layers for feature extraction and a Long Short-Term Memory (LSTM) block to capture temporal dependencies. Using historical glucose, meal, and insulin data, the model was evaluated on the Replace-BG and DIAdvisor datasets, demonstrating strong performance with low errors and accurate glucose-error grid predictions across 30-, 60-, and 90-min horizons. Additionally, QingXiang Bian et al.¹¹ developed a hybrid model that leverages CGM data, using clinical data to combine the Transformer's self-attention mechanism for capturing global dependencies with LSTM's ability to model local patterns. The model achieved mean squared errors (MSE) of 1.18, 1.70, and 2.00 at 15-, 30-, and 45-min intervals, respectively, highlighting its potential for effective blood glucose forecasting.

Existing research in blood glucose prediction faces several challenges, including limited datasets, small sample sizes, and difficulties in capturing long-term dependencies in glucose dynamics. Additionally, many models struggle with noise sensitivity even in high-quality data, have high computational demands, and exhibit slow real-time prediction responses. Few studies attempt to predict blood glucose levels for extended horizons (e.g., 90 or 120 min), due to the increased complexity and uncertainty associated with longer-term forecasting. Moreover, the RMSE values in many models remain high, likely due to short data collection periods that fail to fully capture the complexities of glucose fluctuations.

In this study, we address these limitations by employing a hybrid architecture that combines two deep learning algorithms—Transformer and LSTM—to predict blood glucose levels up to 120 min (30, 60, 90, and 120 min). The Transformer Encoder captures long-range dependencies in the input sequence, generating a contextual representation that enhances the model’s understanding of the entire sequence. Meanwhile, the LSTM processes historical glucose data to capture short-term temporal patterns, allowing the model to leverage both global and local dependencies for better prediction accuracy. However, the direct combination of Transformer and LSTM may struggle to extract higher-level dynamic features in more complex time-series scenarios, limiting its potential to improve prediction accuracy. To address these issues, our model integrates Bidirectional LSTM and Transformer Encoder layers at multiple stages, enabling better joint modeling of dependencies.

To enhance model performance, we incorporate positional encoding and dropout layers while utilizing a sliding window technique to better handle temporal dependencies and reduce noise. Additionally, we enrich the feature set with meal composition and insulin dosage, allowing for more precise prediction of blood glucose fluctuations and minimizing prediction errors. Although our approach does not fully resolve all challenges, it effectively addresses key issues such as overfitting and large prediction errors over extended time periods. Ultimately, our model is designed to predict blood glucose levels for up to 120 min using real-world clinical data, offering clinically valuable insights and demonstrating strong performance in accuracy evaluations through the error grid analysis (EGA) method.

Methods

In this study, we propose an innovative approach that integrates Transformer Encoder and LSTM architectures for blood glucose time series prediction. This approach combines the strengths of both models to address the limitations of using them individually. The research, conducted from May 2024 to the present, is an exploratory and applied study carried out at the Department of Biomedical Engineering, Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan Province.

While both Transformer and LSTM models have been applied to glucose prediction tasks, our method distinguishes itself by employing these models in parallel to capture both global and local dependencies simultaneously. The self-attention mechanism in the Transformer Encoder captures long-range dependencies and global patterns, while LSTM is focused on modeling short-term fluctuations. This parallel approach allows our model to comprehensively address both short-term fluctuations and long-term trends in glucose dynamics—an area where models relying solely on either Transformer or LSTM have shown limitations. In contrast

to previous studies, which process features sequentially through these models, our method simultaneously operates bidirectional LSTM and Transformer Encoder layers, allowing for richer feature abstraction and a better understanding of complex glucose patterns.

Our model further differentiates itself by incorporating additional features that improve prediction accuracy and robustness. Personalized patient features, time-related features, and positional encodings are integrated to enhance the model’s ability to handle dynamic glucose data. These features are processed through different encoding layers and then fused, which enables the model to effectively capture complex temporal dynamics. Additionally, our method includes richer information, such as meal composition and insulin dosage, which is often overlooked by other models that mainly rely on historical glucose data. By combining short-term and long-term dependencies and utilizing a multistage feature fusion mechanism.

Methodology

The proposed model’s input consists of a multivariate time series comprising the patient’s CGM readings along with corresponding timestamps, personalized baseline information (age, gender, height, weight, BMI, waist circumference, hip circumference), meal information for each meal (carbohydrate intake, protein intake, fat intake, calorie intake), and insulin infusion information (insulin dose and basal insulin dose for the four hours prior to infusion) representing the patient’s physiological state. In this scenario, the input data X_t is represented as follows:

$$X_t = [X_{t-L+1}, X_{t-L+2}, X_{t-L+3}, X_{t-L+4}, X_{t-L+5}, \dots, X_{t+L}] \in R^{d \times L} \quad (1)$$

Among them, $X_t \in R^{d \times 1}$ represents d features at time step t , and L is the window length, that is, the number of input time steps.

The focus of this study is to develop a deep learning-based prediction approach, where the model M considers the current input data X_t to predict the future BGC (G_{t+PH}) at time $t + PH$, where prediction horizon (PH) is 30, 60, 90, and 120 min:

$$G_{t+PH} = M(X_t) \quad (2)$$

Transformer model

The application of the Transformer model in time series prediction benefits from its self-attention mechanism and multihead attention mechanism, enabling the capture of long-term dependencies and complex temporal dynamics within sequences.⁶ The self-attention mechanism dynamically allocates attention weights between different time steps, adapting to the importance of different time points in the

sequence to capture long-term dependencies. The multihead attention mechanism allows the model to simultaneously focus on dynamic features at different time scales, facilitating a more comprehensive understanding of the temporal dynamics in the blood glucose sequence.⁶

The self-attention mechanism is defined as follows:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (3)$$

In the formula, Q, K, and V, respectively, represent the Query, Key, and Value tensors, which are all linear transformations of the same input matrix X:

$$Q = XW^Q \quad (4)$$

$$K = XW^K \quad (5)$$

$$V = XW^V \quad (6)$$

The Attention mechanism does not directly use X, but rather employs three matrices generated through matrix multiplication with X. This approach enhances the model's fitting capacity by utilizing three trainable parameter matrices. Q and K^T undergo a Matrix Multiplication operation to produce a similarity matrix, where each element of the similarity matrix is divided by $\sqrt{d_k}$, with d_k representing the dimensionality of K.

The definition of the multihead attention mechanism is as follows:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\alpha_1, \alpha_2, \alpha_3, \alpha_h) * W^O \quad (7)$$

The Concat() function concatenates the outputs of all attention heads together, and W^O represents the final output's linear transformation matrix. In the multihead attention mechanism, the attention weights calculated by each attention head are typically denoted as α_i , where i represents the i -th attention head.

Transformer uses positional encoding to embed position information in the sequence and can take into account the sequential relationship in the time series without the need for recurrent structures. This flexible model architecture can be adapted to various prediction tasks.

Long short-term memory model

Long short-term memory network is a variant of RNNs, which addresses the issue of long-term dependencies and gradient vanishing encountered in traditional RNNs by introducing a gating mechanism.¹² The basic unit of the LSTM network consists of one or multiple memory cells housed in a memory block, along with three adaptive multiplication gate units shared by all cells in the block: an input gate, a forget gate, and an output gate, accompanied by a cell state. These gate units enable the LSTM network to selectively retain or forget past information, facilitating

more effective handling of long sequence data, commonly employed for processing and predicting time series data.

The LSTM input gate controls how the input information at the current time step influences the state of the unit, defined as follows:

$$i_t = \sigma(w_{ii}x_t + b_{ii} + w_{hi}h_{t-1} + b_{hi}) \quad (8)$$

The LSTM forget gate determines which information from previous memory should be ignored in the input and is defined as follows:

$$f_t = \sigma(w_{if}x_t + b_{if} + w_{hf}h_{t-1} + b_{hf}) \quad (9)$$

The LSTM output gate is responsible for controlling how the current hidden state influences the output of the unit and is defined as follows:

$$o_t = \sigma(w_{io}x_t + b_{io} + w_{ho}h_{t-1} + b_{ho}) \quad (10)$$

g_t represents the candidate memory cell and is defined as follows:

$$g_t = \tanh(w_{ig}x_t + b_{ig} + w_{hg}h_{t-1} + b_{hg}) \quad (11)$$

c_t represents the memory cell at the current time step, where \odot denotes the element-wise multiplication (Hadamard product):

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (12)$$

h_t is the hidden state at the current time step, which is based on the output of the memory cell, defined as follows:

$$h_t = o_t \odot \tanh(c_t) \quad (13)$$

g_t is the candidate memory cell, primarily responsible for capturing the relationship between the current input x_t and the previous hidden state h_{t-1} , providing a potential memory update value. It is combined with the input gate i_t through element-wise multiplication, ultimately influencing the content update of the memory cell.

Previously, single LSTM layers have been investigated for blood glucose prediction,¹³ but they lack the ability to fully capture the time dependency exceeding 60 min PH.¹⁴

In the blood glucose prediction task, LSTM can accept past blood glucose values as input and learn the relationship between sequence data to predict future blood glucose values. By training the LSTM network, it can gradually learn and capture patterns and trends in blood glucose data to predict future blood glucose values.

Proposed deep learning model architecture

In the initial experiments, the study evaluated the performance of pure LSTM and pure Transformer models in blood glucose time series prediction tasks. The results revealed that the LSTM model excelled in handling short-term dependencies

but showed limitations in capturing long-term dependencies. Conversely, the pure Transformer model demonstrated strong capabilities in capturing global information and long-range dependencies but was less effective in addressing local dependencies compared to LSTM. Based on these observations, a decision was made to combine the Transformer Encoder with LSTM to leverage the Transformer's ability to capture global information and LSTM's strength in handling local dependencies.

During the architectural design process, multiple experiments were conducted to explore different model combinations and hyperparameter configurations. The results indicated that placing the LSTM layer before the Transformer Encoder effectively captured both short-term and long-term dependencies in the time series. Additionally, experiments were conducted by feeding the data into multiple encoder layers to better extract complex temporal dynamic features in the blood glucose data, which proved to be an effective performance enhancement strategy.

The final architecture design was selected based on a comprehensive consideration of these experimental results and task requirements, aiming to balance the model's complexity with performance. By combining the strengths of LSTM and Transformer Encoder, the model can handle both local and global dependencies simultaneously, thereby improving the accuracy of blood glucose prediction.

In this model architecture, the Transformer Encoder and LSTM are combined, utilizing the Transformer's self-attention mechanism to capture global information and long-range dependencies, while using LSTM to handle short-term dependencies, thereby enhancing prediction accuracy and robustness.

The model's inputs include personalized patient features, time features (converted into sine and cosine values for year, month, day, hour, minute, and second), positional encodings, and blood glucose-related data (such as blood glucose values, insulin doses, and meal data). These features are initially processed through different encoding layers: personalized features and blood glucose-related data are encoded into dense vectors through fully connected layers, time features are transformed into higher dimensions through one-dimensional convolutional layers, and positional encodings are generated by applying sine and cosine functions to sequence positions, enabling the model to perceive relative positional relationships in the input sequence. Subsequently, after all features are converted to the same dimension through different transformation methods, they are fused by direct addition. Specifically, each feature (such as personalized features, time features, location encoding, and blood sugar-related data) is independently transformed and encoded to ensure that their dimensions are consistent. Then, these processed features are fused by element-by-element addition while maintaining the consistency of dimensions, so that the model can process information from different sources simultaneously at each time step.

The core architecture of the model follows a multistage data flow, combining parallel Bidirectional LSTM and Transformer Encoder layers to capture both short-term and long-term dependencies in blood glucose prediction.

First, the fused input features, which have been processed through different transformation methods, are fed separately into the initial Bidirectional LSTM and Transformer Encoder layers (LSTM1 and Encoder1). The LSTM1 is responsible for capturing local temporal dependencies, while Encoder1 extracts global relationships using its self-attention mechanism. The outputs from both LSTM1 and Encoder1 are then passed through linear layers for dimensionality reduction, ensuring they are compatible with the next processing steps.

Next, the output from Encoder1 is passed into the second Transformer Encoder layer (Encoder2), where further global dependencies are captured. Simultaneously, the output from LSTM1 is sent to the second Bidirectional LSTM layer (LSTM2), which focuses on extracting additional short-term and long-range temporal dynamics.

The outputs from LSTM1 and LSTM2 are then fused through element-wise addition, preserving their dimensional consistency. This fusion allows the model to integrate information from both short-term (LSTM1) and long-term (LSTM2) dependencies. After the fusion step, the concatenated features are passed through a linear layer to adjust their dimensionality before they are passed to the final LSTM layer.

Finally, the processed features are passed through the final LSTM layer, which consolidates the information before generating the final prediction. The output from this LSTM layer is then passed through a fully connected layer to produce the final blood glucose predictions. For detailed information regarding the specific hyperparameters used in this model, please refer to Table 1.

Hyperparameter tuning and model optimization

Extensive hyperparameter tuning has been performed before obtaining the proposed optimal model for blood glucose prediction. For example, adjust the number of LSTM layers, the number of units per layer, different activation functions, the number of Transformer encoding layers, the number of heads in the encoding layer, etc. to get the best model. The CGM and insulin data summary are shown in Table 2.

Data sets

Clinical datasets. In this research, the dataset by T1D patients came from the First People's Hospital of Yunnan Province. These data encompassed the BGC of patients at different time points as well as their dietary intake. The CGM device used in the hospital is manufactured by Yunnan Shenyun Maple Technology Co., Ltd, with the

Table 1. Optimal hyperparameter configuration table.

Hyperparameter		Optimal value
Data embedding layer	Linear_indiv_Embedding	Input_dim = 8, Out_dim = 32
	Conv1d_time_Embedding	kernel = 3, stride = 1, padding = 1, dilation = 1, out_channels = 32
	Linear_main_Embedding	Out_dim = 32
	Dim_redu_Linear	Input_dim = 32 * 3, Out_dim = 32
Model	BLSTM1	Hidden_dim = 32, bidirectional = True, num_layers = 2
	Encoder1	d_model = 128, nhead = 4, num_layers = 2
	BLSTM2	Hidden_dim = 64, num_layers = 2
	Encoder2	d_model = 128, nhead = 4, num_layers = 2
	LSTM3	Hidden_dim = 128, num_layers = 2
	Linear1	In_dim = 128, Out_dim = 32
	Linear2	In_dim = 32, Out_dim = 1
Other parameters	Learning rate	0.0001
	Batch size	128
	Windows size	48
	PH	6, 12, 15, 16

model GS1; the insulin pump is manufactured by Dana Diabecare RS, developed by Xuyi Development Co., Ltd, in South Korea. All data collection procedures were conducted in accordance with relevant laws, regulations, and ethical standards and were approved by the hospital's ethics committee. Throughout the data collection process, strict adherence to principles of patient privacy protection was maintained, and confidentiality measures were implemented for the personal information of patients. Written informed consent was obtained from all subjects before the study began. The dataset possesses the following key characteristics:

- The data comprises blood glucose readings, insulin infusion records, and meal intake data for three meals per day across four consecutive weeks for 13 patients. Additionally, personalized baseline data for each patient are included.
- Meal intake data include meal times, carbohydrate intake, protein intake, fat intake, and calorie intake for each meal. Baseline personalized data for

patients include age, gender, height, weight, BMI, waist circumference, and hip circumference.

- Data were obtained from 13 T1D patients (3 males, 10 females), all of whom were undergoing insulin pump therapy.
- A sampling frequency of 5 min was employed to capture blood glucose fluctuations effectively.

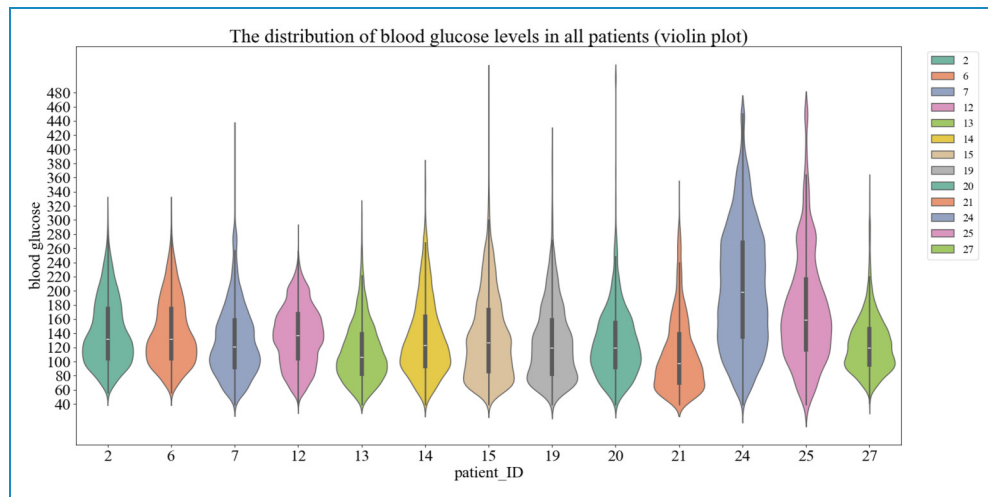
The hyperparameter search space and optimal values are summarized in Table 1.

The violin plot depicting the distribution of patient blood glucose data is shown in Figure 1.

Simulated data set. The UVa/Padova simulator is the only U.S. Food and Drug Administration-approved insulin trial simulator that provides robust and reliable results.¹⁷ The simulator was used to generate 360-day blood glucose data samples of 10 adult T1D patients, with 172,800 blood glucose data instances for each subject. The historical blood glucose values in this data set were accurately sampled every 3 min, making it very suitable to use as input to the model. Each instance contains four data fields: sampling

Table 2. Summary of CGM and insulin data.

ID	Number of test days	Length of the input CGM data	Average glucose level	Glucose std	Number of missing glucose data	Premeal insulin records	4-hour basal dose	Dose miss values count
2	15	4021	142.2	46.882	0	43	49	0
6	15	4021	142.2	46.882	0	42	41	4
7	30	8026	129.6	51.301	0	135	118	0
12	29	8042	131.4	43.203	0	74	78	13
13	44	11910	113.4	42.221	8	135	122	0
14	29	8042	133.2	53.540	0	77	57	10
15	48	12599	136.8	61.708	0	155	132	0
19	15	4385	127.8	54.632	0	48	49	0
20	21	4579	127.8	52.021	0	91	90	0
21	34	8368	109.8	52.687	0	83	83	19
24	39	12063	205.2	85.836	0	55	68	0
25	17	3169	174.6	79.530	0	50	50	1
27	15	4021	124.2	39.839	0	40	42	5

**Figure 1.** Violin diagram of blood glucose data distribution for each patient.

time, CGM value, dietary carbohydrate intake, and insulin dose. The data of each subject is divided into a training set and a test set, with the training set accounting for 80% (138,240 instances) and the test set accounting for 20% (34,560 instances).

Data preprocessing

Missing value imputation and data smoothing. For patient with ID 13 in the clinical dataset, the missing blood glucose data may be due to device malfunctions or abnormal fluctuations caused by individual physiological changes.

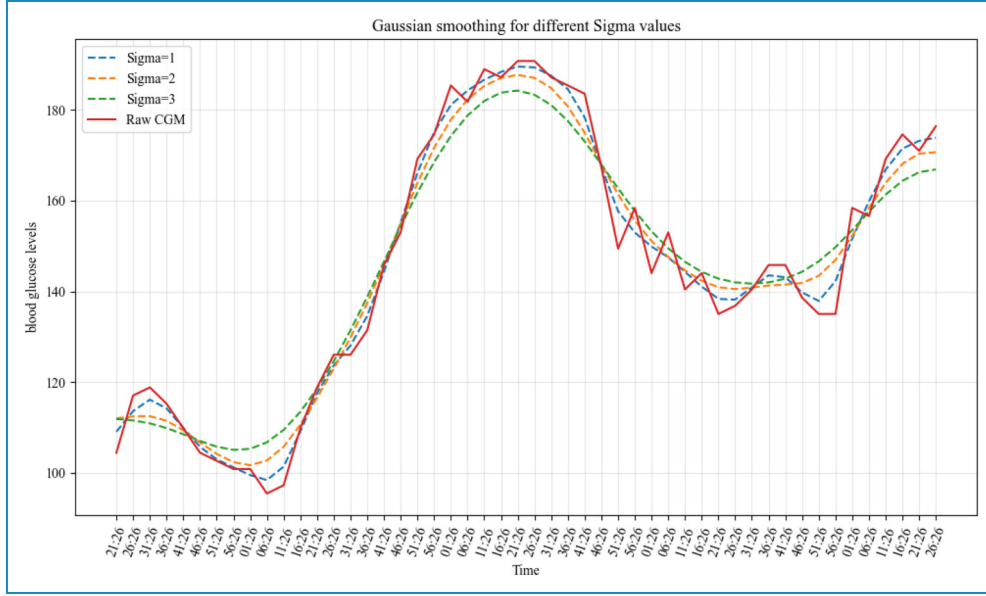


Figure 2. Gaussian smooth comparison of different sigma values.

While the blood glucose data for other patients is complete, there are varying degrees of missing values for meal calories, carbohydrate intake, fat intake, protein intake, premeal insulin dosage, and basal insulin dosage. Considering that the Transformer model used in this study relies on the differences between current and future data points, such discontinuities in the time series may negatively impact the model performance. Furthermore, the proportion of missing blood glucose data points for patient 13 is relatively small (less than 1% of the total data points). Therefore, linear interpolation is chosen to fill in the missing data for patient 13. Additionally, incomplete insulin dosages exist for each patient. Given the daily periodicity of basal insulin dosage, missing values can be filled using data from the previous day.¹⁸

In the clinical dataset, blood glucose data often exhibit high-frequency fluctuations. Continuous glucose monitoring data typically show a sawtooth pattern due to sensor noise, physiological variations, and changes in sampling frequency, necessitating the use of smoothing techniques to enhance data accuracy and usability.¹⁹ Employing one-dimensional Gaussian filtering (G) aids in obtaining a smooth time series:¹⁵

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (14)$$

Simultaneously, it can help improve data quality, reduce noise, and enhance the accuracy and reliability of subsequent analysis and prediction. However, excessively smoothed time series data may remove low and high blood glucose points, leading to significant prediction errors.

Unfiltered continuous blood glucose monitoring trajectories may generate false alarms due to unexpected sensor malfunctions. The degree of filtering needs to strike a balance between accuracy and acceptable prediction errors. In the clinical dataset, the sigma value is set to 1 to maintain an appropriate smoothing effect and ensure that the data still accurately reflects the characteristics of the original blood glucose sequence. The filtering effects of different sigma values on blood glucose time series are illustrated in Figure 2.

Time feature extraction. Converting temporal information into features such as year, month, day, week, hour, and minute enables the model to discern patterns in time series data more effectively. By transforming these temporal features into radians and subsequently applying sine/cosine transformations, time information can be encoded as continuous periodic signals.²⁰ This approach aids the model in comprehending the periodic fluctuations within time series data more accurately and facilitates the capture of continuous temporal variations. Consequently, the model can represent temporal information as abstract features rather than mere specific dates or timestamps.

Data normalization. When handling personalized patient feature data, the first step involves normalizing the features to the [0,1] range. This step aims to eliminate the differences in scales between features, ensuring that they are within the same numerical range. This mitigates the risk of certain features disproportionately influencing model training, thereby enhancing model stability and reliability. Normalized data also accelerate the convergence speed

of the model and reduce its susceptibility to noise interference.²¹

Window partitioning and dataset splitting. The data provided by the hospital include blood glucose data, insulin bolus doses, and meal records for 13 T1D patients. Initially, the blood glucose, insulin, and meal data are aligned based on time to ensure completeness at each time point. Subsequently, a sliding window approach is applied to partition the data sequences for each patient. The window length is set to 48 with a step size of 1. The input features for each window consist of a 48-length time series, with the target value being the blood glucose values at n time steps after the window ends. In this process, data from past time steps are used as inputs, while data from future time steps are used as target values, transforming the time series prediction problem into a standard supervised learning problem.

Specifically, each sample consists of a 48-length input sequence and a n -length label sequence, denoted as $([x_1, x_2, x_3, \dots, x_{48}], [y_1, y_2, \dots, y_n])$. For a blood glucose value sequence of length L , it can be divided into $L - 48 - n + 1$ samples, where n represents the prediction range. Subsequently, all samples are split into training and testing sets in a ratio of 8:2.

Main_feat represents the main features, Time_feat represents the time features (sine and cosine transformations of year, month, day, hour, minute, and second), and BaseInf represents the personalized features. The neural network model for blood glucose prediction is illustrated in Figure 3.

Evaluation indicators

Analysis and evaluation. Based on previous research, in this study RMSE and mean absolute error (MAE) are selected as evaluation indicators. The RMSE calculation formula is:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (15)$$

where n is the number of samples; y_i is the actual observed value of the i th sample; \hat{y}_i is the predicted value of the i th sample; Σ denotes summation over all samples.

The MAE calculation formula is:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (16)$$

where n , y_i , \hat{y}_i , and Σ are the same as in the RMSE formula.

Clinical evaluation. Error grid analysis is widely recognized as a clinical interpretation in blood glucose prediction tasks.²² Therefore, in this study EGA is used to evaluate the severity of errors in clinical aspects. The measured BGC and predicted values are depicted on a scatter plot

with five significant regions. Based on these assumptions, the grid is divided into five different regions representing varying degrees of accuracy and inaccuracy in blood glucose estimation, as illustrated in Figure 4.

Zone A represents the range where the deviation between predicted values and reference values does not exceed 20%, or when the reference value is below 70 mg/dL, the blood glucose value falls within the hypoglycemic range (<70 mg/dL). Blood glucose values in this range are considered clinically accurate as they aid in making appropriate treatment decisions.²²

Upper and lower parts of Zone B represent values deviating from the reference by $>20\%$, which, according to assumptions, would result in benign treatment or no treatment.²²

Zone C values would lead to excessive correction of acceptable blood glucose levels, potentially causing actual blood glucose to fall below 70 mg/dL or rise above 180 mg/dL.²²

Zone D represents errors of “dangerous failure to detect and treat.” Actual blood glucose values are outside the target range, but the patient-generated blood glucose values fall within the target range.²²

Zone E is the “mistreatment” zone. Values generated by the patient in this area are opposite to the reference values, thus resulting in treatment decisions opposite to those required. In summary, values in Zones A and B are clinically acceptable, while those in Zones C, D, and E pose potential risks and are therefore significant clinical errors.²²

Results

Experimental setup

Both the clinical dataset and the simulated dataset consist of two independent files for model training and testing. The training set is used for model construction, training, and hyperparameter tuning. After each training epoch, the model undergoes validation on the validation set. This validation process includes computing the total validation loss and the model’s prediction accuracy on the validation set. Performance evaluation metrics are calculated based on the predicted results obtained from the test set. The model undergoes five training runs, each comprising 300 epochs, utilizing early stopping with a patience value of 10. The best-performing model on the validation set is saved after each run.

The initial learning rate is set to 0.0001. Adam optimizer is chosen, which adapts the learning rate for each parameter based on the first and second moments of the gradients, leading to more efficient parameter updates and accelerated convergence. Additionally, different parameters have different learning rates, enhancing adaptability.²³ During each iteration on the training set, the model simultaneously processes 128 samples and updates its parameters based on

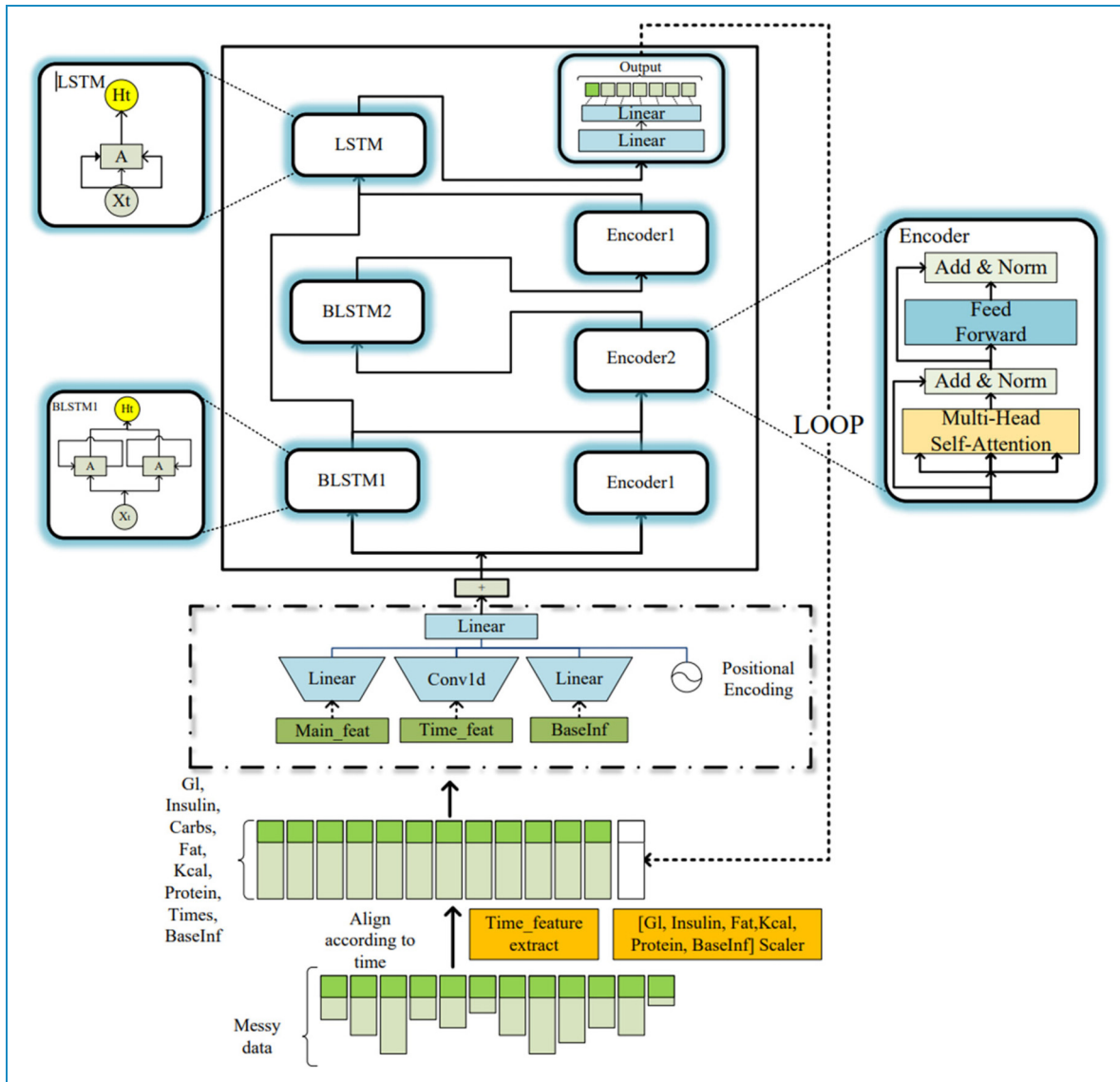


Figure 3. Diagram of the proposed deep learning model.

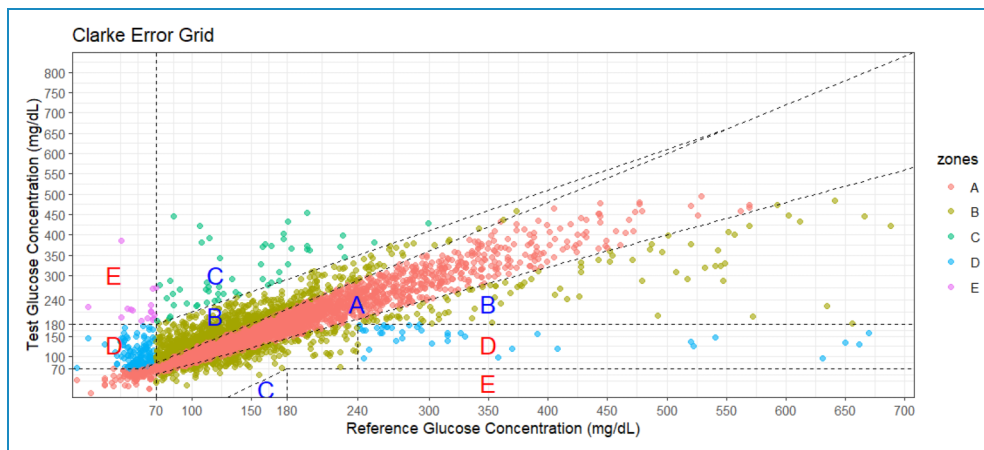


Figure 4. Error grid analysis diagram example.

Table 3. RMSE and MAE with different prediction ranges for each patient.

PID	30-min		60-min		90-min		120-min	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	5.302	3.607	7.682	4.088	8.169	4.24	10.349	4.746
6	6.038	3.895	5.945	3.931	6.854	4.264	8.545	4.572
7	9.955	8.45	10.389	8.928	11.788	8.953	13.086	8.317
12	7.109	4.738	8.03	5.015	9.608	5.467	22.332	6.746
13	21.482	13.121	22.206	12.722	21.924	12.764	19.803	11.359
14	6.238	4.319	8.064	4.954	9.361	5.967	12.396	6.809
15	11.764	10.336	11.643	9.428	12.642	10.015	14.971	10.648
19	9.685	6.942	10.19	7.461	11.804	7.791	14.116	7.883
20	12.213	8.053	12.639	8.007	14.474	7.746	20.736	9.935
21	12.223	8.531	12.891	8.744	16.43	9.456	17.157	10.013
24	4.018	2.816	5.27	3.071	8.093	3.808	12.822	4.276
25	6.412	3.945	6.044	4.033	9.128	5.159	10.622	5.233
27	5.439	5.439	5.241	3.522	9.508	4.782	9.114	4.538

the average loss of these samples. Python 3.10 is used as the programming language, and PyTorch 2.0.0 + cu117 framework is employed for model construction and training. PyTorch 2.0.0 + cu117 is a stable version of the PyTorch deep learning framework built on CUDA 11.7, enabling efficient GPU utilization for computational acceleration. The hardware resources utilized for training and performance evaluation consist of a computer equipped with an Intel Core i5 13490F processor and Nvidia GeForce RTX 3060Ti graphics card, with 16GB of memory.

Experimental results

Comparison of prediction ranges for different patients. Table 3 presents the RMSE and MAE of the model for each patient on the test set. Each column displays the RMSE and MAE at different PHs. Comparatively, the prediction performance is better for a 30-min PH than for extended PHs. This behavior is evident as increasing the PH adds complexity. The decrease in prediction performance resulting from the expansion of PHs may also be attributed to increased opportunities for activities such as exercise, insulin dosing, and/or meals.²⁴

As the PH increases, errors generally grow due to accumulated inaccuracies, leading to instability in long-term predictions. Patients 13 and 20 in Table 3 show the highest errors, mainly due to discontinuities in their blood glucose data, where gaps between data points disrupt the linear interpolation process.

Surprisingly, the RMSE for patient 13 decreases with longer horizons (e.g., 120 min, RMSE = 19.803). This could be because the model captures the overall blood glucose trend better over longer periods, smoothing out short-term fluctuations and improving prediction accuracy.

The model's Clarke EGA plots for different PHs on the test set are shown in Figure 5. As the PH values increase, the predicted results become more dispersed. However, it is encouraging to note that even when the PH value reaches 120 min, the majority of predictions still fall within the clinically accepted boundaries of zones A and B. Table 4 summarizes the clinical assessment of the EGA on the clinical dataset, indicating that the model's performance remains at an acceptable level even under higher PH conditions.

Additionally, for qualitative assessment, Figure 6 presents the 24-h blood glucose predictions and reference trajectories for patient ID 12 at 30, 60, 90, and 120 min PHs, respectively. The cyan solid line represents the reference

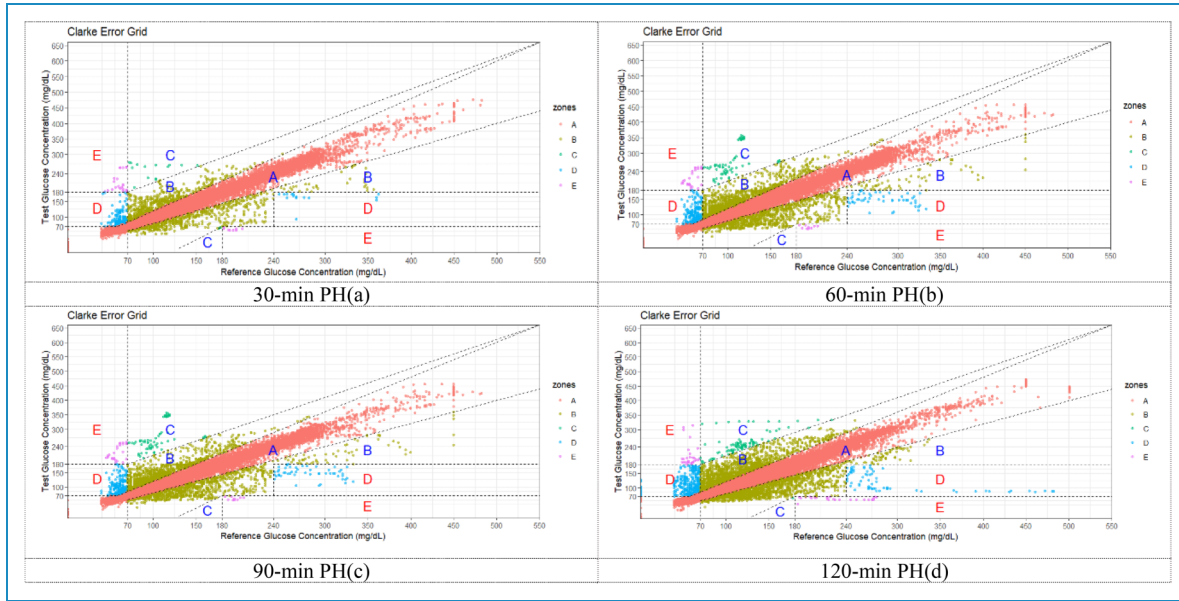


Figure 5. Error grid analysis (EGA) plots for different prediction ranges on the test set.

Table 4. Summary of Clarke error grid analysis for different PH.

PH	Region A + B (%)	Region C + D + E (%)
30-min	99.597	0.403
60-min	99.017	0.982
90-min	97.562	2.437
120-min	96.366	3.63

Note. “A + B” represents the combination of regions A and B, representing clinical safe predictions. “C + D + E” represents the combination of regions C, D, and E, collectively representing clinically unsafe predictions.

BGC values obtained from CGM devices, while the red solid line denotes the predictions made by the proposed model.

Comparison with other methods. Through the same procedure, the simulator-generated data underwent preprocessing and integration before being input into our model for prediction. Table 5 presents the prediction errors (RMSE and MAE) for different PHs using the UVa/Padova simulator dataset, comparing our model with others. The results of all these comparison methods are taken from the corresponding papers, and all models used the same dataset. Our model (Ls-Encoder) outperforms others with significantly lower RMSE and MAE across all horizons, particularly at 30 and 60 min. Overall, the Ls-Encoder model demonstrates superior performance in both short- and

medium-term predictions, reducing errors and surpassing most existing methods.

Table 6 summarizes the performance of different models in blood glucose prediction on the clinical dataset, comparing RMSE and MAE metrics. For the 30-min prediction, Gluformer achieved the best performance (RMSE 8.01, MAE 6.20), while Ls-Encoder showed slightly inferior results (RMSE 10.157, MAE 8.87). In the 60-min prediction, Ls-Encoder demonstrated competitive performance close to Gluformer (RMSE 10.645 vs. 10.39, MAE 10.11 vs. 9.75). For the 90-min and 120-min predictions, Ls-Encoder exhibited significant advantages, achieving the best performance in the 120-min prediction (RMSE 13.986, MAE 6.986). Overall, Ls-Encoder outperformed other models in long-term predictions.

Discussion

In this study, our proposed model exhibited strong performance in blood glucose prediction tasks, particularly in the long-term prediction scenarios (90 min and 120 min), where it outperformed other recent approaches. Error grid analysis results also show that our model performs well in clinical accuracy, providing predictions within clinically acceptable ranges. The error generally increased with the extension of the prediction range, a typical challenge in long-term glucose prediction due to the accumulation of inaccuracies over time.

When comparing the results of our model with other recently proposed methods (as shown in Table 6), our hybrid model, which integrates Bidirectional LSTM and Transformer Encoder layers in parallel stands out by

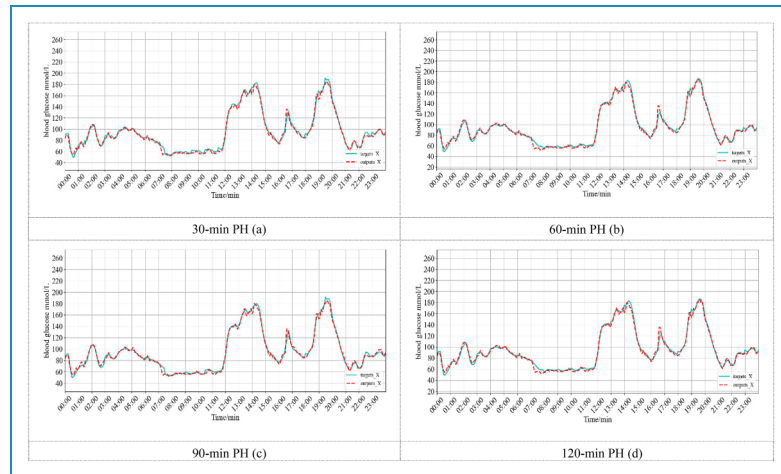


Figure 6. Comparison of predicted values and reference values in different prediction ranges for patient 12 in one day. (a) The deviation between the predicted values and the actual values is relatively small. (b) The deviation between the predicted values and the actual values begins to increase. (c) The deviation between the predicted values and the actual values becomes more pronounced. (d) The deviation between the predicted values and the actual values is the greatest.

Table 5. Summary of prediction errors for different models at various pH levels in the simulated dataset.

PH	No	Related literature	YOP	Model used	RMSE (mg/dL)	MAE (mg/dL)
15 min	1	Ignacio et al. ¹⁶	2019	ARIMA	10.15	—
	2	Mohebbi et al. ²⁵	2020	LSTM	11.68	7.90
	3	Kalita et al. ²⁶	2022	LS-GRUNet	5.27	2.91
	4	Saúl Langerica et al. ²⁷	2023	MetaL	3.78	—
	5	This work		Ls-Encoder	1.793	1.376
30 min	6	Alessandro Aliberti et al. ¹³	2019	LSTM	19.47	—
	7	Hadia Hameed et al. ²⁸	2020	Vanila RNN	19.21	13.08
	8	Taiyu Zhu et al. ²⁹	2021	Edge LSTM	19.10	13.59
	9	Kalita et al. ²⁶	2022	LS-GRUNet	14.58	11.04
	10	Saúl Langerica et al. ²⁷	2023	MetaL	5.73	—
	12	This work		Ls-Encoder	2.049	1.311
60 min	13	Mohebbi et al. ²⁵	2019	LSTM	36.57	26.85
	14	Bhimi Reddy et al. ³⁷	2020	LSTM + BiLSTM	42.46	—
	15	Cichosz et al. ³⁰	2021	LOCF	41.53	—
	16	Kalita et al. ²⁶	2022	LS-GRUNet	35.04	23.54
	17	Saúl Langerica et al. ²⁷	2023	MetaL	7.19	—
	18	This work		Ls-Encoder	3.477	1.668

Note. Bold values indicate the best performance algorithm for each PH.

Table 6. Summary of prediction errors for different models at various pH levels in clinical dataset.

PH	No	Related literature	YOP	Model used	RMSE (mg/dL)	MAE (mg/dL)
30 min	1	Albertetti Fabrizio et al. ²⁸	2020	BLGP-HES-SO	10.42	7.12
	2	Ran Cui et al. ³¹	2021	GluPred	11.63	9.09
	3	Renat Sergazinov et al. ³⁸	2023	Gluformer	8.01	6.20
	4	This work		Ls-Encoder	10.157	6.377
60 min	5	Albertetti Fabrizio et al. ²⁸	2020	BLGP-HES-SO	13.89	10.31
	6	Ran Cui et al. ³¹	2021	GluPred	12.84	11.96
	7	Renat Sergazinov et al. ³⁸	2023	Gluformer	10.39	9.75
	8	This work		Ls-Encoder	10.645	6.417
90 min	9	Albertetti Fabrizio et al. ²⁸	2020	BLGP-HES-SO	17.18	13.67
	10	Ran Cui et al. ³¹	2021	GluPred	16.26	12.52
	11	Renat Sergazinov et al. ³⁸	2023	Gluformer	13.87	8.06
	12	This work		Ls-Encoder	13.537	7.283
120 min	13	Albertetti Fabrizio et al. ²⁸	2020	BLGP-HES-SO	22.08	18.41
	14	Ran Cui et al. ³¹	2021	GluPred	20.31	14.52
	15	Renat Sergazinov et al. ³⁸	2023	Gluformer	15.43	9.85
	16	This work		Ls-Encoder	13.986	6.986

simultaneously capturing both short-term and long-term dependencies. Existing models often face limitations, such as reliance on short data durations, computational complexity, and sensitivity to data quality. For instance, approaches such as ANN and RNN models are constrained by short-term data, limiting their ability to capture long-term glucose trends. Additionally, models that separate LSTM and Transformer layers often lose valuable information when processing features individually. These methods also tend to overfit, especially when trained on small datasets or noisy data.

Our hybrid model integrates Bidirectional LSTM and Transformer Encoder layers in parallel, enabling the simultaneous capture of both short- and long-term dependencies, which improves prediction accuracy. By combining richer features, such as personalized patient data and meal information, and incorporating dropout layers to prevent overfitting, our model is more robust and adaptable, even in the presence of noisy or incomplete data. Experimental results demonstrate that this innovative integration and enriched feature set significantly enhance the accuracy and robustness of

blood glucose prediction, providing a more comprehensive and reliable framework for long-term glucose forecasting compared to other state-of-the-art methods.

In parallel, by incorporating glucose data, insulin dosages, and meal composition, our model can help predict and manage diabetes onset or worsening in post-COVID patients, supporting timely clinical interventions.³² Integrating metabolomic data could also improve the early identification of patients at risk for severe outcomes, enhancing personalized T1D management.³³ Our model analyzes real-time glucose data and clinical indicators to provide personalized predictions, helping mitigate risks similar to those in hypertension-related brain damage.³⁴ Additionally, incorporating immune response parameters can enhance predictions of T1D patients' responses to immunotherapy, optimizing treatment plans.³⁵ Finally, while some studies focus on insulin transcription, our model offers a long-term perspective by predicting blood glucose levels and insulin-glucose interactions.³⁶

Despite promising results, the internal mechanisms and decision-making processes of the model remain unclear.

In clinical practice, model interpretability is crucial for gaining the trust of healthcare professionals and patients. Although this study demonstrated a favorable performance, future research must prioritize enhancing the model's transparency to provide clearer insights into how predictions are made. Additionally, the dataset used in this study was limited to T1D patients from specific regions, which could restrict the generalizability of the model. Future research should focus on diversifying the dataset by including data from different regions and ethnic groups, which would enhance the model's applicability to a broader population.

Moreover, while the current model has demonstrated its potential in laboratory settings, further optimization is necessary to improve its performance and efficiency. Input from clinical experts is crucial for developing practical models that can be integrated into real-world clinical practices.

In conclusion, this study demonstrates the promising potential of combining Transformer and LSTM models for blood glucose prediction in T1D patients. The findings serve as a solid foundation for future research aimed at improving diabetes management. However, it is critical to recognize the current limitations, including the generalizability of the model, the interpretability of its predictions, and the potential biases introduced by dataset limitations. Future work should focus on optimizing the model's structure, expanding the dataset, and collaborating with clinical experts to refine the model's real-world applicability. Ultimately, these efforts will help improve the quality of life for T1D patients by providing more accurate, personalized diabetes management solutions.






Acknowledgements

This work was supported by Special Project for “Famous Doctor” of Yunnan Ten Thousand Talents Plan (grant numbers YNWR-MY-2019-020).

Guarantor

XX

ORCID iDs

Xin Xiong  <https://orcid.org/0000-0002-2206-5206>
 XinLiang Yang  <https://orcid.org/0009-0005-8013-677X>
 Yunying Cai  <https://orcid.org/0000-0003-2355-1502>
 Yuxin Xue  <https://orcid.org/0009-0008-7618-7218>
 JianFeng He  <https://orcid.org/0009-0002-4766-6718>

Statements and declarations

Ethical approval

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Yunnan First People's Medical Ethics Committee (protocol code KHLL2022-KY165, and date of approval: 18 October 2022).

Informed consent

Informed consent was obtained from all subjects involved in the study.

Contributorship

Conceptualization, Xin Xiong and JianFeng He; methodology, Xin Xiong and Yunying Cai; software, Yuxin Xue; validation, Xin Xiong, Yunying Cai, and JianFeng He; formal analysis, Xin Xiong and JianFeng He; investigation, Xin Xiong and JianFeng He; resources, XinLiang Yang; data curation, XinLiang Yang; writing—original draft preparation, Yunying Cai; writing—review and editing, Yuxin Xue and JianFeng He; visualization, Heng Su; supervision, JianFeng He; project administration, XinLiang Yang; funding acquisition, XinLiang Yang. All authors have read and agreed to the published version of the manuscript.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability

The datasets generated and/or analyzed during the current study include two datasets: 1. The dataset collected from T1D patients in the First Renmin Hospital of Yunnan Province is not publicly available due to privacy concerns. 2. The second dataset was generated using the UVa/Padova simulator. This dataset is not directly available, but interested researchers can generate similar data using the UVa/Padova simulator, which can be downloaded from [<https://tegvirginia.com/t1dms/>].

References

1. Kotagal M, Symons RG, Hirsch IB, et al. Perioperative hyperglycemia and risk of adverse events among patients with and without diabetes. *Ann Surg* 2015; 261: 97–103.
2. Cryer PE, Davis SN and Shamoon H. Hypoglycemia in diabetes. *Diabetes Care* 2003; 26: 1902–1912.
3. Ramachandran A, Ananth S and Chiranjeevi MP. Deep learning based time series modelling for glucose level prediction of type-1 diabetes. 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2023: 1–6.
4. Bremer T and Gough DA. Is blood glucose predictable from previous values? A solicitation for data. *Diabetes* 1999; 48: 445–451.
5. Zhu T, Li K, Chen J, et al. Dilated recurrent neural networks for glucose forecasting in type 1 diabetes. *J Healthcare Inf Res* 2020; 4: 308–324.
6. Lee SM, Kim DY and Woo J. Glucose transformer: forecasting glucose level and events of hyperglycemia and hypoglycemia. *IEEE J Biomed Health Inform* 2023; 27: 1600–1611.

7. Bertachi A, Biagi L, Contreras I, et al. Prediction of blood glucose levels and nocturnal hypoglycemia using physiological models and artificial neural networks. *KHD@IJCAI* 2018; pp. 85–90.
8. Cappon G, Meneghetti L, Prendin F, et al. A personalized and interpretable deep learning based approach to predict blood glucose concentration in type 1 diabetes. *KDH@ECAI* 2020; 20: 75–79.
9. Li K, Liu C, Zhu T, et al. Glunet: a deep learning framework for accurate glucose forecasting. *IEEE J Biomed Health Inform* 2019; 24: 414–423.
10. Long-Term Prediction of Blood Glucose Levels in Type 1 Diabetes Using a CNN LSTM-Based Deep Neural Network.
11. Bian QX, As' arry A, Cong XG, et al. A hybrid transformer-LSTM model apply to glucose prediction. *PLoS One* 2024; 19: e0310084.
12. Hochreiter S and Schmidhuber J. Long short-term memory. *Neural Comput* 1997; 9: 1735–1780.
13. Aliberti A, Pupillo I, Tema S, et al. A multi-patient data-driven approach to blood glucose prediction. *IEEE Access* 2019; 7: 69311–69325.
14. Hermans M and Schrauwen B. Training and analysing deep recurrent neural networks. *Adv Neural Inf Process Syst* 2013; 26.
15. Li K, Daniels J, Liu C, et al. Convolutional recurrent neural networks for glucose prediction. *IEEE J Biomed Health Inform* 2019; 24: 603–613.
16. Rodríguez-Rodríguez I, Chatzigiannakis I, Rodríguez JV, et al. Utility of big data in predicting short-term blood glucose levels in type 1 diabetes mellitus through machine learning techniques. *Sensors* 2019; 19: 4482.
17. Herrero P, Georgiou P, Oliver N, et al. A bio-inspired glucose controller based on pancreatic β -cell physiology. *J Diabetes Sci Technol* 2012; 6: 606–616.
18. Yang T, Wu R, Tao R, et al. Multi-Scale long short-term memory network with multi-lag structure for blood glucose prediction. *KDH@ECAI* 2020; 45: 136–140.
19. Johnson E and White F. Understanding the physiological basis of blood glucose fluctuations in type 1 diabetes patients. *Diabetes Care* 2017; 40: 456–462.
20. Salinas D, Flunkert V, Gasthaus J, et al. DeepAR: probabilistic forecasting with autoregressive recurrent networks. *Int J Forecast* 2020; 36: 1181–1191.
21. Ioffe S and Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning*. PMLR 2015; pp. 448–456.
22. Clarke WL, Cox D, Gonder-Frederick LA, et al. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care* 1987; 10: 622–628.
23. Kingma D P and Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
24. Shuvo MMH and Islam SK. Deep multitask learning by stacked long short-term memory for predicting personalized blood glucose concentration. *IEEE J Biomed Health Inform* 2023; 27: 1612–1623.
25. Mohebbi A, Johansen AR, Hansen N, et al. Short term blood glucose prediction based on continuous glucose monitoring data//2020. 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020; pp. 5140–5145.
26. Kalita D and Mirza K B. LS-GRUNet: glucose forecasting using deep learning for closed-loop diabetes management. 2022 IEEE 7th International conference for Convergence in Technology (I2CT), 2022; pp. 1–6.
27. Langarica S, Rodríguez-Fernandez M, Núñez F, et al. A meta-learning approach to personalized blood glucose prediction in type 1 diabetes. *Control Eng Pract* 2023; 135: 105498.
28. Freiburghaus J, Rizzotti A and Albertetti F. A deep learning approach for blood glucose prediction of type 1 diabetes. *Proceedings of the Proceedings of the 5th International Workshop on Knowledge Discovery in Healthcare Data Co-located with 24th European Conference on Artificial Intelligence (ECAI 2020)*, 29-30 August 2020, Santiago de Compostela, Spain. p. 2675.
29. Zhu T, Kuang L, Li K, et al. Blood glucose prediction in type 1. Diabetes using deep learning on the edge. 2021 IEEE International Symposium on Circuits and Systems (ISCAS), 2021; pp. 1–5.
30. Cichosz SL, Kronborg T, Jensen MH, et al. Penalty weighted glucose prediction models could lead to better clinically usage. *Comput Biol Med* 2021; 138: 104865.
31. Cui R, Hettiarachchi C, Nolan CJ, et al. Personalised short-term glucose prediction via recurrent self-attention network. 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), 2021; pp. 154–159.
32. Shen J, Wang Y, Deng X, et al. Combining bioinformatics and machine learning algorithms to identify and analyze shared biomarkers and pathways in COVID-19 convalescence and diabetes mellitus. *Front Endocrinol (Lausanne)* 2023; 14: 1306325.
33. Spagnolo P, Tweddell D, Cela E, et al. Metabolomic signature of pediatric diabetic ketoacidosis: key metabolites, pathways, and panels linked to clinical variables. *Mol Med* 2024; 30: 250.
34. Avvisato R, Forzano I, Varzideh F, et al. A machine learning model identifies a functional connectome signature that predicts blood pressure levels: imaging insights from a large population of 35 882 patients. *Cardiovasc Res* 2023; 119: 1458–1460.
35. Van Rampelbergh J, Achenbach P, Leslie RD, et al. First-in-human, double-blind, randomized phase 1b study of peptide immunotherapy IMCY-0098 in new-onset type 1 diabetes: an exploratory analysis of immune biomarkers. *BMC Med* 2024; 22: 259.

36. Wong WKM, Thorat V, Joglekar MV, et al. Analysis of half a billion datapoints across ten machine-learning algorithms identifies key elements associated with insulin transcription in human pancreatic islet cells. *Front Endocrinol (Lausanne)* 2022; 13: 853863.
37. himireddy A, Sinha P, Oluwalade B, et al. Blood glucose level prediction as time-series modeling using sequence-to-sequence neural networks. *CEUR Workshop Proceedings* 2020.
38. Sergazinov R, Armandpour M and Gaynanova I. Gluformer: transformer-based personalized glucose forecasting with uncertainty quantification. *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023: pp. 1–5.