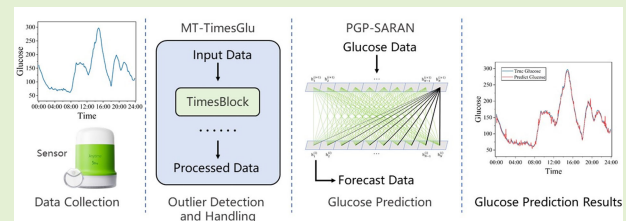


Blood Glucose Anomaly Correction and Personalized Short-Term Prediction Based on Continuous Glucose Monitoring Data

Zhuoyun Jiang¹, Zhichao Yu¹, Suhai Yuan, Yutong Han¹, Guanqun Zhang, and Zhigang Zhu¹

Abstract—Diabetes, as a chronic metabolic disease, relies on precise blood glucose monitoring and management for treatment. The quality of continuous glucose monitoring (CGM) data and prediction accuracy are critical for achieving closed-loop control in an automated insulin delivery system (AIDS). To address challenges such as sensor anomalies and physiological delays in CGM data, this study proposes an attention-based framework for glucose anomaly detection and short-term prediction. First, we design an anomaly detection and correction module that integrates a multiperiod spatiotemporal attention mechanism with a dual-task joint learning framework to simultaneously optimize anomaly identification and interpolation. In addition, a blood glucose prediction module based on a recurrent self-attention mechanism and transfer learning is proposed. Experimental results show that the proposed anomaly detection model achieves an accuracy of 96.4%, with Pearson correlation coefficients (PearsonCCs) and intragroup correlation coefficients between interpolation results and actual values reaching 0.854 and 0.846, respectively. On the Yuwell dataset, the glucose prediction module attains average root mean square errors (RMSEs) of 6.27, 10.40, and 14.31 mg/dL at 15, 30, and 45 min. The average clinical safety zone coverage in Clarke error grid analysis reaches 99.52%. On the public OhioT1DM dataset, the average clinical safety zone coverage reaches 98.92%. Our approach significantly outperforms some existing methods in detection sensitivity, prediction accuracy, and clinical applicability, providing a robust and practical solution for diabetes management.

Index Terms—Continuous glucose monitoring (CGM), glucose prediction, outlier detection, recursive self-attention mechanism, spatio-temporal attention mechanism.



I. INTRODUCTION

DIABETES is a chronic metabolic disease caused by multiple pathogenic factors, the main characteristic of which is that blood sugar levels continue to rise above the normal range [1]. The global diabetes patient population reached

828 million in 2022, an increase of 630 million compared to 1990, with the prevalence rate also showing an upward trend year by year [2]. The prevention and treatment of diabetes has thus become a critical focus in global medical research.

Early strategies for managing diabetes have predominantly focused on intermittent monitoring of blood glucose levels combined with medication-based interventions. Both methods required invasive blood sampling and provided only point-in-time glucose information [3]. In recent years, continuous glucose monitoring (CGM) technology [4] has made significant advancements. By implanting glucose sensors subcutaneously, CGM enables continuous, real-time monitoring of blood glucose levels while reducing patient discomfort. Integrating CGM devices with insulin pumps forms an automated insulin delivery system (AIDS), enabling closed-loop control of blood glucose levels in diabetic patients.

In AIDS, the effectiveness of blood glucose control highly depends on the quality of blood glucose monitoring data. However, in practical use, blood glucose sensors may produce abnormal values due to various reasons, and these abnormal

Received 5 August 2025; accepted 11 August 2025. Date of publication 20 August 2025; date of current version 2 October 2025. This work was supported in part by the University Capacity Building Project of Shanghai Science and Technology Commission under Grant 21010502800. The associate editor coordinating the review of this article and approving it for publication was Dr. Deepak Joshi. (Zhuoyun Jiang and Zhichao Yu contributed equally to this work.) (Corresponding author: Zhigang Zhu.)

Zhuoyun Jiang, Zhichao Yu, Suhai Yuan, Yutong Han, and Zhigang Zhu are with the School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China (e-mail: zyjiang@usst.edu.cn; yuzhch2000@163.com; 243383696@st.usst.edu.cn; yutonghan@usst.edu.cn; zhigang_zhu259@163.com).

Guanqun Zhang is with the Jiangsu Yuwell-POCT Biological Technology Company, Ltd., Danyang, Jiangsu 212300, China (e-mail: zhang.gq1@yuyue.com.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JSEN.2025.3598254>, provided by the authors.

Digital Object Identifier 10.1109/JSEN.2025.3598254

values can negatively impact closed-loop algorithms, thereby affecting blood glucose control outcomes. In addition, since CGM devices measure blood glucose levels by detecting glucose concentrations in subcutaneous tissue, and subcutaneous glucose concentrations lag behind changes in blood glucose levels, there is some delay in the actual measurement results [5]. When used in conjunction with an insulin pump, the injected insulin also requires a certain period of time to take effect after entering the body [6]. Therefore, CGM devices must also have a certain degree of blood glucose prediction capability to eliminate the effects of time delays.

In the field of blood glucose anomaly detection, the scarcity of labeled abnormal data has led to the widespread adoption of unsupervised and semisupervised learning techniques [7]. Turksoy et al. [8] proposed a method combining unscented Kalman filtering for state estimation to detect CGM failures. This method relies on preset physiological dynamics assumptions and is not adaptable enough to nonlinear sudden abnormalities. Yu et al. [9] used a sparse recursive kernel filtering algorithm to improve the reliability of CGM sensors in artificial pancreas systems. However, its reliance on statistical features limits its adaptability to complex physiological changes. A variety of machine learning approaches [10] have been explored, including clustering algorithms, isolation forests, and one-class support vector machines. Zhao et al. [11] developed a semisupervised outlier detection method for CGM data using an independent central point OPTICS clustering algorithm. However, the effectiveness of this approach is highly dependent on the selection of clustering parameters, which may limit its robustness. Li et al. [12] used k -medoids clustering technology for outlier identification, but this method relies on random initialization of k cluster centers, resulting in significant differences in the final clustering results. To address this, Li et al. [12] integrated DBSCAN with a convolutional neural network (CNN) for noninvasive glucose monitoring using ECG signals. However, DBSCAN's effectiveness depends heavily on the careful selection of neighborhood parameters, which typically requires prior knowledge. Idi et al. [13] systematically evaluated the detection performance of unsupervised algorithms such as isolated forests for CGM faults. Although these methods avoid complex modeling, they rely on manual feature design and have limited sensitivity to long-term gradual anomalies. In summary, most unsupervised or clustering methods struggle to achieve satisfactory results in blood glucose anomaly detection. Therefore, this article designs a supervised anomaly detection model, explicitly introduces fault prior information, achieves more accurate anomaly identification and interpolation repair on the labeled dataset, and provides a more stable data basis for subsequent blood glucose prediction.

Blood glucose prediction approaches can be broadly categorized into physiology-based and data-driven models. Physiology-based models simulate glucose–insulin dynamics by incorporating multiple influencing factors, including dietary intake, physical activity, and endogenous insulin response. For example, Bergman et al. [14] defined and calculated the insulin sensitivity index by monitoring the dynamic changes of blood glucose and insulin levels after intravenous glucose injection and analyzing the glucose disappearance rate using a specific

nonlinear formula. Sorensen [15] developed a physiological model of human glucose metabolism that integrates 22 differential equations to simulate glucose–insulin dynamics. Although this approach is highly interpretable, it is computationally intensive. Georga et al. [16] proposed a multivariate prediction model for subcutaneous glucose concentration in patients with type 1 diabetes using support vector regression (SVR). The model integrates multiple physiological inputs to capture individual variability. However, its performance is sensitive to input feature selection and lacks robustness in handling noisy CGM data. Although physiological models have been widely applied in blood glucose prediction and have demonstrated effectiveness in specific scenarios, their reliance on empirically derived parameters limits their ability to generalize across individuals with diverse glycemic regulation mechanisms. In contrast, data-driven approaches, powered by machine learning and data mining techniques, can autonomously learn latent patterns of glucose variation directly from historical data without requiring prior physiological assumptions.

The emergence of data-driven approaches, particularly deep learning methods, has partially overcome these limitations and has gradually become the mainstream in blood glucose prediction due to their strong temporal modeling capabilities. Georga et al. [17] used a kernel adaptive filter and analyzed the impact of multivariate features, including subcutaneous blood glucose, carbohydrate intake, insulin regimen, and physical activity, on the accuracy of short-term glucose prediction in patients with type 1 diabetes. Although the model improved accuracy by integrating multiple sources of physiological inputs, it relied on accurate and complete records of insulin and dietary data, limiting its practicality in real-world scenarios. In real-world scenarios, it is often impractical to continuously obtain comprehensive physiological data (e.g., meal intake, insulin dosage, injection times, and physical activity). The complexity and subjectivity involved in accurately recording these features may introduce noise and reduce the reliability of model inputs.

Several studies have demonstrated that blood glucose prediction based solely on historical CGM data is both feasible and effective. Lu et al. [18] demonstrated that in open-loop settings, short-term glucose forecasting can be effectively performed using only historical CGM readings, without the need to include intricate contextual variables. A related study by Hamdi et al. [19] introduced a blood glucose forecasting approach that combines SVR with a differential evolution algorithm, targeting patients with type 1 diabetes. Their model relies solely on CGM data as input and demonstrates strong predictive accuracy and stability across various forecasting intervals. Therefore, this study adopts a univariate modeling approach using only CGM historical data. This choice not only aligns with practical deployment scenarios but also avoids the complications associated with data fusion from unreliable or unavailable external sources.

In the field of time-series analysis, recurrent neural networks (RNNs) and long short-term memory (LSTM) networks are among the most widely adopted deep learning architectures. RNNs possess powerful nonlinear representation capabilities, enabling them to capture temporal dependencies in sequential

data. The LSTM network, as a special variant of RNN, is able to retain long-term information and effectively learn complex temporal patterns and has been shown to improve the prediction performance of blood glucose prediction tasks. Li et al. [20] proposed a combination of CNN and LSTM to predict the blood glucose levels of patients with type 1 diabetes for the next 30 and 60 min. Experiments show that the model shows high prediction accuracy on simulated and real patient data. Rabby et al. [21] developed a deep RNN model that employs stacked LSTM layers to enhance the capability of modeling long-term dependencies in glucose prediction. By increasing the sequence depth, the model achieved improved temporal representation over standard LSTM structures. However, the unidirectional nature of LSTM limits its ability to extract contextual patterns from both past and future time steps, a limitation that can be addressed by bidirectional architectures such as bidirectional LSTM (Bi-LSTM). Madan et al. [22] proposed a network that combines CNN and Bi-LSTM to improve the real-time blood glucose prediction performance of patients with type 2 diabetes.

Although traditional RNN- and LSTM-based models have achieved success in capturing temporal dependencies, they suffer from inherent limitations. RNNs are prone to gradient vanishing or explosion when modeling long sequences. While LSTMs, despite their gated architecture designed to mitigate these issues, still struggle to effectively capture ultra-long-range dependencies. The transformer model, with its self-attention mechanism and ability to capture global dependencies, has emerged as a powerful tool for time-series forecasting in recent years. Despite its strong global modeling capacity, the standard transformer architecture exhibits limited sensitivity to local signal fluctuations and personalized physiological dynamics, and it incurs high computational costs. Furthermore, existing models often overlook the integration of outlier detection, leading to degraded performance when anomalous data are present. Therefore, we propose an enhanced transformer-based architecture tailored to better accommodate the physiological dynamics and local fluctuations inherent in blood glucose signals. This study focuses on two critical aspects of CGM: robust outlier detection and accurate glucose prediction. The main contributions of this work are summarized as follows.

- 1) We propose an enhanced multitask TimesNet for glucose (MT-TimesGlu) model that integrates a multiperiod spatiotemporal collaborative attention mechanism to improve sensitivity to short-term anomalies. By employing a dual-task joint learning framework, the model simultaneously optimizes outlier detection and imputation, thereby improving the quality and reliability of input data and laying a solid foundation for accurate and personalized blood glucose prediction.
- 2) To address the challenges of long-range dependency modeling and individual variability, we developed a blood glucose prediction model named personalized glucose prediction with self-attentive and recurrent attention network (PGP-SARAN). This model employs a stacked recursive self-attention encoder to capture multilevel temporal features and integrates a global

dynamic attention mechanism to enhance the modeling of physiological rhythms. In addition, it leverages transfer learning to reduce the reliance on large amounts of target-domain data, thereby improving personalized prediction performance under limited sample conditions.

- 3) We conducted extensive experiments to validate the effectiveness of the proposed framework. The evaluation was based on the publicly available OhioT1DM dataset and the Yuwell dataset to ensure the generalizability and reliability of the results. A comprehensive set of performance metrics was employed to quantitatively assess both outlier detection and glucose prediction accuracy, demonstrating the superiority of our proposed approach.

The remainder of this article is organized as follows. Section II provides a detailed description of the proposed outlier detection model. Section III presents the blood glucose prediction model. Section IV discusses the experimental results and analysis. Finally, Section V concludes this article with a summary and discussion.

II. BLOOD GLUCOSE ABNORMALITY DETECTION AND PROCESSING

This study focuses on outlier detection, handling, and blood glucose value prediction in blood glucose monitoring data. First, in response to potential outlier interference in CGM data, we proposed an outlier detection and processing module, MT-TimesGlu, based on the attention mechanism. Subsequently, based on the processed blood glucose data, a recursive self-attention-driven blood glucose prediction module, PGP-SARAN, was designed. The two modules work together to form a complete workflow from anomaly handling to blood glucose prediction, providing reliable data support for diabetes closed-loop management.

This section systematically presents the design and implementation of an attention-based model for blood glucose abnormality detection and correction. First, we define the criteria for identifying abnormal glucose values. Second, we describe the proposed enhanced multitask and multiperiod spatio-temporal attention model, which is designed to jointly address anomaly detection and imputation. The overall architecture of the model is illustrated in Fig. 1.

A. Definition of Outliers

The accuracy of blood glucose prediction depends heavily on the quality of the input data. Outliers are measurements that deviate significantly from the physiological norm (e.g., <2.2 mM or >22.3 mM) or the central tendency of the dataset [23]. The problem may be caused by CGM sensor calibration errors or performance degradation. This study focuses on four common types of CGM-related anomalies, and waveform examples are shown in Fig. 2.

1) *Spike, i.e., Peak Fault*: A sudden, transient deviation in the CGM signal, usually caused by electromagnetic interference or internal sensor noise, appears as a sudden high or low peak, as shown in Fig. 2(a) and is modeled by the following equation:

$$y_e(k) = y(k) + d_{im} y(k). \quad (1)$$

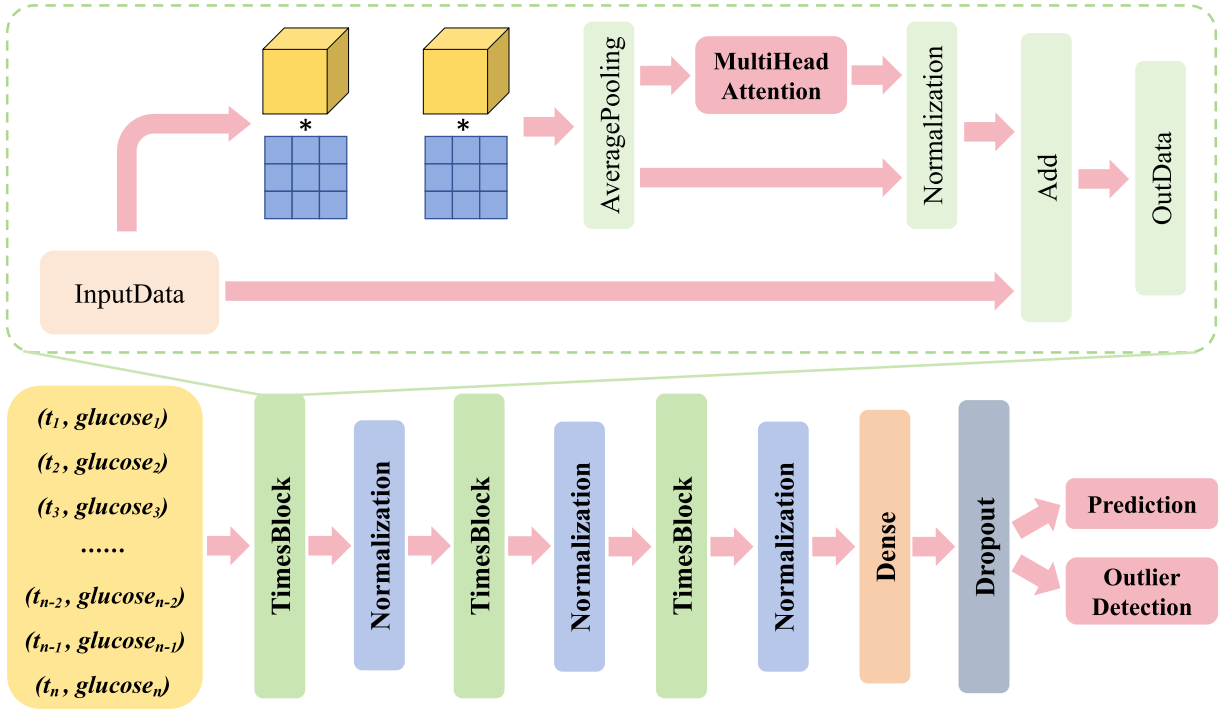


Fig. 1. MT-TimesGlu model structure diagram.

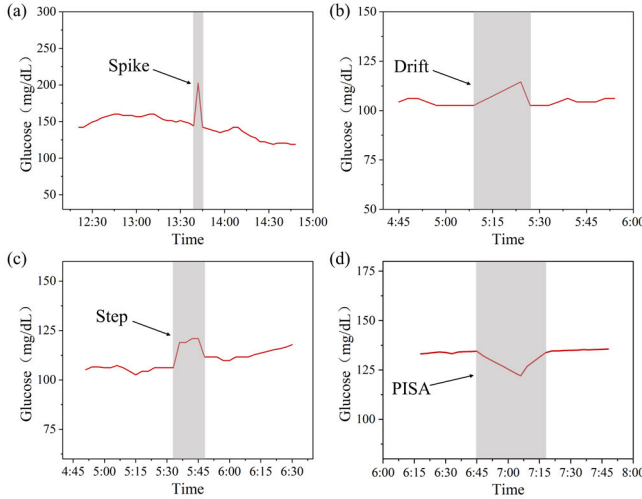


Fig. 2. Schematic of common sensor fault waveforms. (a) Spike fault. (b) Drift fault. (c) Step fault. (d) PISA fault.

Among them, $y_e(k)$ is the generated erroneous CGM measurement value, $y(k)$ is the original CGM value, and $d_i m_a y(k)$ is a random deviation value.

2) *Drift, i.e., Drift Failure*: A gradual deviation of sensor readings from true values over time, typically caused by sensor aging or environmental influences, manifests as a slow upward or downward trend in the CGM waveform, as shown in Fig. 2(b), and is formulated in the following equation:

$$\begin{aligned} & [y_e(k), y_e(k+1), \dots, y_e(k+d_u-1)] \\ &= [y(k), y(k+1), \dots, y(k+d_u-1)] \\ &+ \frac{d_i m_a y(k) [12, \dots, d_u]}{d_u}. \end{aligned} \quad (2)$$

Among them, d_u is an integer randomly selected from the range [2], [3], [4], [5], representing the duration of the drift.

3) *Step, i.e., Step Fault*: Step refers to the phenomenon of sudden step-like changes in sensor measurements. The main causes of step faults include sensor failure and external impact. When a step fault occurs, the corresponding CGM waveform of the patient will show a noticeable step, forming a sudden rise or fall, as shown in Fig. 2(c). Step fault can be expressed using the following equation:

$$\begin{aligned} & [y_e(k), y_e(k+1), \dots, y_e(k+d_u-1)] \\ &= [y_e(k), y_e(k+1), \dots, y_e(k+d_u-1)] \\ &+ d_i m_a y(k). \end{aligned} \quad (3)$$

Among them, d_u is an integer randomly selected from the range [2], [3], [4], [5], representing the duration of the step.

4) *PISA or Pressure-Induced Sensor Attenuation*: PISA refers to the phenomenon of decreased sensitivity or weakened signals caused by continuous pressure on the sensor. When PISA occurs, the corresponding CGM waveform of the patient shows a gradual downward trend, forming a slow concave waveform. After the pressure is relieved, the waveform gradually recovers, as shown in Fig. 2(d). PISA fault can be expressed using the following equation:

$$\begin{aligned} & y_e(k+t) \\ &= \begin{cases} y(k+t) - P \left(1 - \exp \left(\frac{-xt}{\tau} \right) \right), & \text{if } t \leq \frac{u}{5} \\ y(k+t) + P \left(1 - \exp \left(\frac{-xt+u}{\tau} \right) \right) - P \left(1 - \exp \left(\frac{-xt}{\tau} \right) \right), & \text{if } \frac{u}{5} < t < d_u. \end{cases} \end{aligned} \quad (4)$$

Among them, $d_u = (u + 3\tau/5)$, $t \in [1, d_u]$, $\tau \in [5, 10, 15, 20]$, $u \in [15, 20, 25, 30]$, and $P \in [15, 20, 25, 30]$.

B. TimesNet

TimesNet [24] is a deep learning framework for time series modeling. The model is designed to capture multi-period temporal patterns through a modular structure that decomposes complex time series into periodic components. By transforming 1-D time series into a 2-D representation, TimesNet enables unified modeling of intra- and interperiod variations. Its core module, TimesBlock, consists of four main steps: 1-D-to-2-D transformation, 2-D periodic pattern extraction, 2-D-to-1-D reconstruction, and adaptive fusion. This architecture effectively models periodic dynamics across different time scales and enhances the expressive power of the network in capturing rich temporal dependencies.

C. Multitask and Multiperiod Spatio-Temporal Attention Model

TimesNet is a temporal data foundation model whose primary feature is the expansion of 1-D temporal data into a 2-D space for analysis. However, it suffers from high computational costs and a high degree of dependence on data periodicity. Blood glucose temporal data exhibits characteristics such as nonstationarity and noise interference. Applying TimesNet to blood glucose temporal modeling faces challenges such as insufficient computational power and the inability to simultaneously handle outlier detection and missing value handling. To address these issues, this study designed MT-TimesGlu, which integrates a multiperiod spatiotemporal collaborative attention mechanism with a dual-task joint learning framework. This model can detect blood glucose anomalies while filling in missing values. Its main components are detailed as follows.

1) *Multiperiod Spatio-Temporal Attention Mechanism*: Traditional time-based models often fail to capture sudden fluctuations in blood glucose levels caused by meals, physical activity, or sleep cycles. To address this issue, we designed a new TimesBlock based on physiological knowledge of blood glucose fluctuations, introducing a multiperiod spatio-temporal attention structure (as shown in Fig. 1) to effectively simulate short-term changes, postmeal cycles, and circadian rhythms. This enhances the model's ability to detect sudden changes in blood glucose levels caused by external factors. First, the input sequence is transformed into a 2-D tensor via a reshape operation according to a predefined cycle p , covering short-term fluctuations, postmeal cycles, and circadian rhythms. Local spatio-temporal features are extracted using cascaded 2-D convolutions, as shown in (5). Subsequently, a multihead attention mechanism is employed to capture long-range dependencies across time cycles. Finally, feature fusion is achieved through layer normalization and residual connections. Through these improvements, the MT-TimesGlu model's sensitivity to short-term anomaly detection is significantly enhanced, while the model's training convergence speed is further accelerated

$$\mathbf{X}_{\text{conv}} = \text{Conv2D}(\text{ReLU}(\text{Conv2D}(\mathbf{X}_{\text{reshaped}}))) \quad (5)$$

where $\mathbf{X}_{\text{reshaped}}$ is the input 2-D tensor; Conv2D is the 2-D convolution operation; ReLU is the activation function used to introduce nonlinearity and set negative values to 0; and $\mathbf{X}_{\text{conv}}$ is the final output.

2) *Dual-Task Joint Learning Framework*: In traditional outlier detection and handling tasks, outlier detection and missing value imputation are typically treated as independent, sequential tasks. This approach has some limitations, such as the inability of single-task frameworks to capture the relationship between outlier events and adjacent data missingness and the inability to leverage the correlation between multitask objectives to enhance the representational power of shared features. To address these limitations, we propose a dual-task joint learning framework based on parameter sharing and dynamic gradient balancing, which unifies anomaly detection and value interpolation under a shared feature and coordinated optimization process.

The MT-TimesGlu model constructs a shared feature extractor by stacking improved TimesBlock blocks. The input sequence is processed through a spatio-temporal collaborative attention mechanism to generate high-dimensional representations. The model includes two independent task decoders: the regression branch outputs the corrected results of abnormal blood glucose values through a fully connected layer, while the detection branch outputs the probability score of anomaly detection using a sigmoid activation function, enabling simultaneous anomaly detection and outlier interpolation of blood glucose values. By designing a weighted joint loss function for training, the model addresses gradient conflicts in multitask learning and determines the optimal weight ratio through grid search.

3) *Period Presetting Based on Clinical Prior Knowledge*: In blood glucose time series data analysis, most models use traditional FFT period detection methods to identify the primary periods in the time series. However, these methods are highly sensitive to noise, leading to invalid period identification and affecting detection performance. They may also generate non-physiological periods that clinical doctors cannot explain or verify. This study designed a manual period presetting scheme based on clinical prior knowledge to match the physiological periodic patterns of blood glucose changes. For example, a 2-h window is associated with the typical postprandial blood glucose metabolic response period, a 5-h window may reflect the complete postprandial blood glucose metabolic cycle, and a 12-h window is used to capture the circadian rhythm changes in blood glucose [25], [26]. Compared to FFT cycle detection, this approach significantly reduces the time required for cycle detection, avoids the computational overhead of dynamic calculations, and improves model training speed while maintaining good cycle modeling capability. It also reduces spurious harmonics and enhances the physiological plausibility of the learned patterns.

4) *Model Lightweight Design*: In real-world medical scenarios, patient blood glucose data are limited, and using complex models can lead to overfitting. In this study, while ensuring model performance, the number of stacked attention modules was reduced, and the convolutional structure was simplified, reducing the total number of parameters from 2.35 to 0.87 M.

This reduces both the number of parameters and computational complexity while improving the model's generalization performance. In the model parameter section, in the temporal attention module, the periodic parameter period is set to [2], [5], and [10]; in the convolutional layer, filters is set to 64, kernel_size is set to 3, and a double-layer Conv1D is used to extract local features, with the activation function being ReLU; in the attention layer, the number of attention heads is 4, and the dimension of each head is 64. In the training parameter section, the batch size is 32, the total number of training epochs is 100, the initial learning rate is set to 0.001, and the learning rate optimizer is the Adam optimizer. When the training loss does not improve within 3 epochs, the learning rate is reduced to 50% of the original value. The model's loss function employs a joint loss function comprising two tasks, where the loss functions for predicting corrected blood glucose values and anomaly detection probability scores are shown in the following equations:

$$L_{\text{prediction}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i). \quad (6)$$

$$L_{\text{anomaly_score}} = \frac{1}{N} \sum_{i=1}^N (s_i - \hat{s}_i). \quad (7)$$

Among them, y_i is the true value, $\hat{y}_i$ is the corrected value, N is the sample size; s_i is the true anomaly score, and $\hat{s}_i$ is the anomaly score predicted by the model. The joint loss is shown in the following equation:

$$L = \alpha \cdot L_{\text{prediction}} + \beta \cdot L_{\text{anomaly_score}} \quad (8)$$

where α and β are the weights for the prediction and anomaly detection tasks, respectively. An early stopping mechanism was used in the experiment. When the loss function did not improve within 10 epochs, training was stopped, and the model weights with the smallest loss were restored.

Rather than treating anomaly detection and correction as isolated processes, our method integrates them within a unified multitask framework, enabling joint optimization and mutual enhancement. Furthermore, by introducing a multiperiod spatio-temporal attention mechanism, the model captures both short-term fluctuations and long-range dependencies more effectively. This design significantly improves predictive performance over traditional single-task or unidirectional approaches. The detailed procedure of the proposed method is outlined in Algorithm 1.

III. BLOOD GLUCOSE PREDICTION MODEL

This study proposes a personalized blood glucose prediction model, PGP-SARAN [as shown in Fig. 3(a)], based on a recursive self-attention mechanism. By utilizing a hierarchical recursive self-attention architecture and a dynamic weight allocation mechanism, the model enhances its ability to model blood glucose time-series features. Compared to traditional time series models, PGP-SARAN can better capture long-term dependencies and local fluctuations in blood glucose levels even with limited training data. The main components of PGP-SARAN are detailed below.

Algorithm 1 Blood Glucose Anomaly Detection and Imputation

Input: dataset $G = \{g_1, g_2, g_3, \dots, g_m\}$; the window size w ; the threshold percentile
Output: Outlier $G_{\text{out}} = \{g_1, \dots, g_n\}$; Imputed value $I_{\text{value}} = \{i_1, \dots, i_n\}$

- 1: Load CGM data G
- 2: Preprocess data: normalize glucose values
- 3: Create time series windows of size w
- 4: Construct the TimesNet-based detection model:
- 5: Input layer: TimesBlock(period = 2)
- 6: TimesBlock(period = 5)
- 7: TimesBlock(period = 10)
- 8: Dense layers
- 9: Dual output (prediction + anomaly
- 10: score)
- 11: Train model with Adam optimizer and callbacks
- 12: Perform anomaly detection:
- 13: Calculate reconstruction error
- 14: Set threshold at given percentile
- 15: Mark points exceeding threshold as outliers
- 16: Impute faulty values based on anomaly type (step/spike/drift/pisa):
- 17: for each fault point:
- 18: if index < w : use previous value
- 19: else:
- 20: predict value with TimesBlock
- 21: Save outliers and imputation results after processing
- // Sub-algorithm: Times Block
- Input: 3D tensor (batch, seq, features)
- 22: Reshape to 4D tensor (batch, period, seq/period, features)
- 23: Apply double Conv2D layers
- 24: Global average pooling $\rightarrow$ RepeatVector
- 25: Apply MultiHeadAttention with residual connection
- 26: Layer normalization $\rightarrow$ Output

A. Temporal Series Features

Traditional RNN or LSTM models struggle to capture long-range dependencies in blood glucose fluctuations due to gradient vanishing and limited local receptive fields. To address this issue, this study introduces a multihead self-attention mechanism [27], which uses a dynamic weight matrix to model global dependencies across time steps. Specifically, the model adaptively allocates attention weights in the input sequence, prioritizing nearby time points while capturing global dependencies, thereby effectively enhancing its ability to capture long-range dependencies. Fig. 3(b) illustrates how the self-attention mechanism captures long-term dependencies and key time points. The top and bottom rows in the figure represent the hidden states of the current layer and the previous layer, respectively. Each hidden state $h_j^{(i+1)}$ of the current layer is dynamically associated with all time-step hidden states $\{h_1^{(i)}, h_1^{(i)}, \dots, h_N^{(i)}\}$ of the previous layer through the self-attention mechanism. The attention weight matrix adaptively assigns weights based on the importance of each time step in

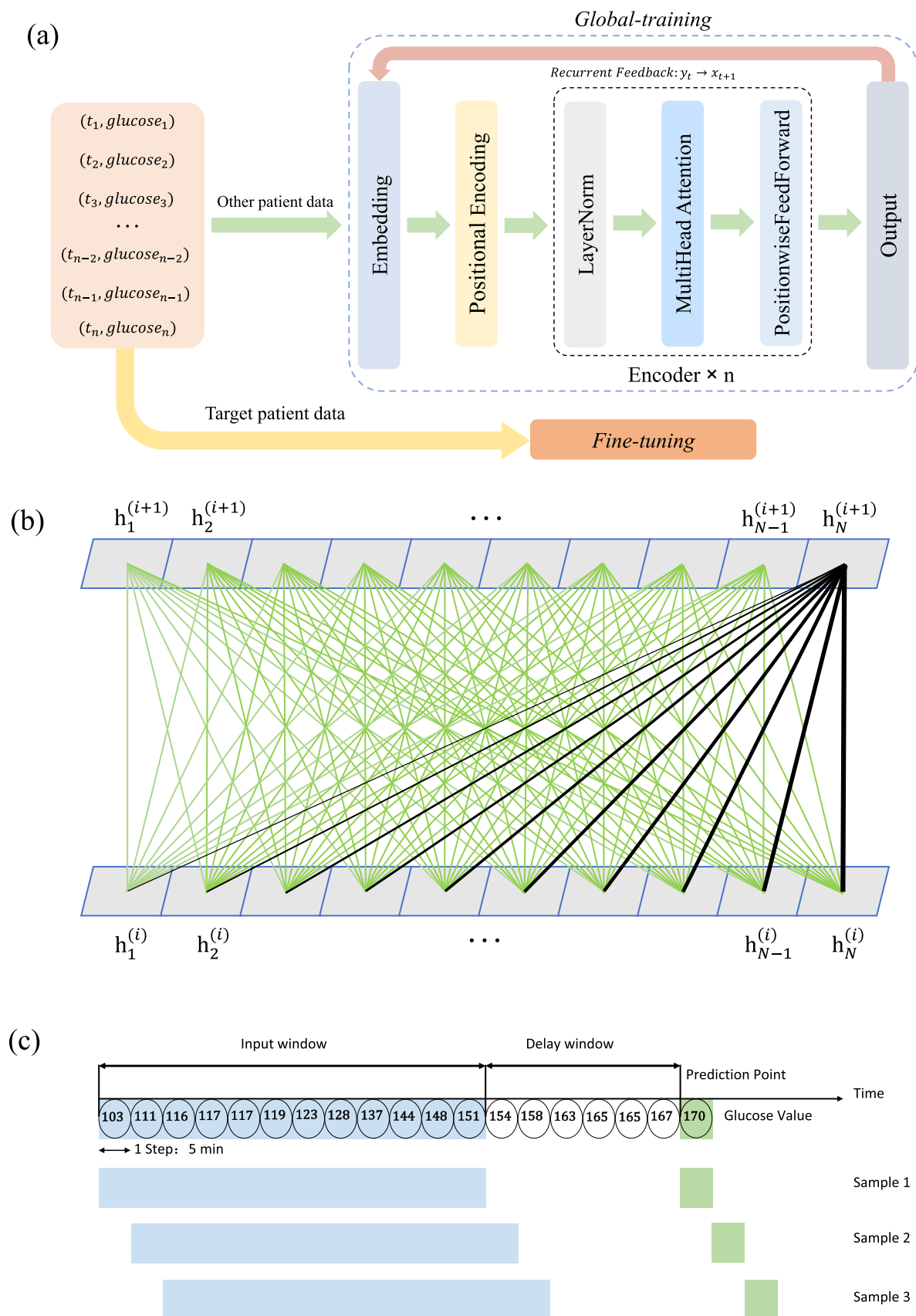


Fig. 3. Blood glucose prediction model. (a) Schematic of the PGP-SARAN model structure. (b) Schematic of the dynamic temporal attention mechanism. (c) Schematic of blood glucose concentration prediction process.

the input sequence, thereby enhancing focus on neighboring time points. The thickness of the lines in the figure reflects the level of weight assignment, with thicker lines indicating higher weights. By stacking multiple self-attention layers, the model can extract temporal features layer by layer, with each layer further enhancing its ability to model long-range dependencies, thereby effectively capturing global dependencies in blood glucose fluctuations.

To model global dependencies among historical glucose readings, we first apply a multihead self-attention mechanism to the extracted feature sequence. The attention (Q, K, V) for each layer can be calculated as

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (9)$$

where Q, K , and V represent the query, key, and value matrices, respectively. These are derived from the hidden representations T via linear transformations using projection weights W^Q, W^K , and W^V .

Multihead attention applies several parallel attention heads to capture diverse dependencies. Each head computes attention independently, and the outputs are concatenated and linearly projected to form the final result. The computation is given by the following equation:

$$\begin{aligned} \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h) W^o \\ \text{head}_i &= \text{Attention}\left(Q W_i^Q, K W_i^K, V W_i^V\right) \end{aligned} \quad (10)$$

where $W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W^o \in \mathbb{R}^{hd_v \times d_{\text{model}}}$, h is the number of attention heads, d_{model} is the dimension of the input vector, and $d_v = d_{\text{model}}/h$. This attention mechanism enables the model to flexibly aggregate relevant temporal information from both near and distant positions, thereby enhancing its ability to learn complex, long-range dependencies in blood glucose dynamics.

B. Recursive Self-Attention

While multihead attention provides the capability to extract global dependencies, it alone may not be sufficient to model complex temporal dynamics and fine-grained local variations in blood glucose sequences. To further enhance the model's ability to capture multiscale temporal features, we introduce a recursive self-attention architecture. First, multiple self-attention layers are stacked to extract features at different scales. Each layer includes residual connections and layer normalization to mitigate gradient vanishing or explosion issues, ensuring stable model training. Second, a local attention mechanism is introduced in some network layers, enabling the model to better capture short-term fluctuations in blood glucose values while maintaining focus on the overall time series trend. Finally, a sine-based position encoding method is used to ensure the continuity of temporal information, addressing the potential temporal disorder issues that traditional transformers may encounter in physiological signals, thereby ensuring the model can accurately capture temporal information.

In terms of training methods, this study also employs transfer learning techniques, with the model training process divided into two stages: global training and fine-tuning training. During the global training stage, the model is trained using data from nontarget patients to grasp the common characteristics of the diabetic patient population, laying the foundation for subsequent personalized predictions. During the fine-tuning stage, the model is trained using personalized data from target patients to adjust model parameters to adapt to individual characteristics and patterns. Through transfer learning technology, the model can converge quickly on a small amount of personal data, improving prediction performance and reducing the need for a large amount of personal data.

C. Sliding Window Strategy for Glucose Prediction

To achieve real-time blood glucose prediction, we adopted a sliding window method to simulate the behavior of a CGM system in a real-world environment. As shown in Fig. 3(c), each prediction cycle consists of three key components: the input window, the delay window, and the prediction point. The input window contains a fixed-length sequence of historical blood glucose values, with the input window size set to 12 time steps in the figure, each separated by a 5-min interval; the delay window represents the prediction delay window, with a size of 6 in the figure, corresponding to a delay time of 30 min, and the prediction point is the target prediction point at the current time step. After completing the prediction target for the current time step, the system waits for the CGM to transmit the actual value for the next time step. Once the data are received, the input window, delay window, and target prediction point all shift backward by one time step. Based on the new input window, the system completes the prediction for the new target prediction point, and this process repeats in a loop to complete the prediction for the entire blood glucose sequence.

In actual CGM use, the overall delay is relatively large due to the slow change in interstitial fluid glucose concentration and the delayed action of insulin. Therefore, the prediction delay window of the prediction model generally needs to be greater than 30 min to be of practical application value [28]. In the experiment, the size of the historical data window input for prediction was 2 h, to predict the patient's blood glucose levels 15-, 30-, and 45 min in the future. The detailed procedure of the proposed method is outlined in Algorithm 2.

IV. EXPERIMENTS RESULTS AND ANALYSIS

A. Datasets and Preprocessing

In the outlier detection and handling experiment, the experimental data used the Yuwell dataset provided by Jiangsu Yuwell-POCT Biological Technology Company Ltd. This dataset includes CGM data from 20 real diabetic patients. The dataset includes glucose levels and fingerstick blood glucose information, with a monitoring period of 14 days. For each patient's blood glucose data, we divided it into a training set and a test set in an 8:2 ratio. The patient information in the dataset is shown in Supplementary Table S-I.

To evaluate the robustness of the proposed anomaly detection framework, four typical types of faults—step, spike,

Algorithm 2 Blood Glucose Prediction Based on Recursive Self-Attention

Input: CGM data G , Known sequence length L_{left} , Prediction horizon $L_{missing}$

Output: Predicted glucose values $\hat{G}$ for $L_{missing}$ future time steps

Data Preprocessing and Model Initialization

1: Detect and fill missing glucose values using linear interpolation

2: Create fixed-length sequences:
sequence_length = $L_{left} + L_{missing}$

3: Initialize Transformer encoder with parameters

Autoregressive Prediction

4: For each sequence:

5: Initialize context window: context = $X[0 : L_{left}]$

6: Apply multi-head self-attention to context

7: Generate glucose prediction: $g_t =$

8: Transformer(context)

9: Update context: context[t] = g_t

10: Output predictions: $\hat{G} = [g_{L_{left}} \dots g_{L_{left}+L_{missing}-1}]$

Two-Stage Training

11: Global pre-training phase: Train on other patient dataset, Optimize using MSE loss

12: Personal fine-tuning phase: Fine-tune on target patient data, Use reduced learning rate, Early stopping based on validation RMSE

Evaluation

13: Calculate performance metrics: RMSE, MAE and MARD

drift, and PISA—were artificially introduced into the original Yuwell CGM dataset, as defined by (1)–(4). These faults were randomly injected into the time series to simulate real-world abnormalities frequently encountered in CGM systems. By incorporating these synthetic faults, the experiments aim to comprehensively assess the model's ability to accurately detect and process different types of anomalies under controlled conditions.

In the blood glucose prediction experiment, we used both the Yuwell dataset and the OhioT1DM dataset. The OhioT1DM dataset was released by Ohio University and includes blood glucose data from 12 real T1DM patients. The dataset contains glucose levels (glucose level), finger stick, bolus insulin, basal insulin, temp. basal, and meal information. The OhioT1DM dataset contains 8 weeks of blood glucose data from patients, divided into training and testing sets [29]. Patient information in the OhioT1DM dataset is shown in Supplementary Table S-II.

B. Performance Metrics

In outlier detection tasks, TP represents the number of outliers correctly detected, FP represents the number of normal values incorrectly detected as outliers, FN represents the number of outliers incorrectly detected as normal values, and TN represents the number of normal values correctly detected. In real-world data, the number of normal values far exceeds

the number of outliers. Even if the model predicts all samples as normal values, the accuracy rate may still be high, but this does not necessarily indicate that the model possesses actual detection capability. Therefore, when evaluating the performance of outlier detection algorithms, we primarily focus on the success rate of detecting outliers, i.e., the proportion of correctly detected outliers among all actual outliers

$$NF = TP + FN \quad (11)$$

$$S = \frac{TP}{NF} \quad (12)$$

among them, S is the success detection rate.

In the evaluation of outlier handling, evaluation metrics are selected from two perspectives: correlation and consistency. In the correlation evaluation, the Pearson correlation coefficient (PearsonCC) is used as the evaluation metric. In consistency evaluation, the intraclass correlation coefficient (ICC) is used as the evaluation metric. PearsonCC is used to measure the linear correlation between interpolated values and true values, while ICC is used to assess the statistical stability and reproducibility of interpolated values relative to true values. A PearsonCC value closer to 1 or -1 is preferable, while an ICC value closer to 1 is preferable.

The equation for calculating Pearson CC is as follows:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (13)$$

In the blood glucose prediction task, we use the root mean square error (RMSE), mean absolute error (MAE), and mean absolute relative error (MARD) as error metrics to evaluate the overall predictive performance of the model.

RMSE primarily assesses the model's generalization and robustness, MAE is more suitable for evaluating prediction performance during stable blood glucose periods, and MARD directly reflects the algorithm's prediction accuracy. Smaller values for these three metrics indicate better model performance. The equations for calculating the three metrics are as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (14)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (15)$$

$$MARD = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (16)$$

In the above equation, y_i represents the actual value, $\hat{y}_i$ represents the predicted value, and N represents the total number of samples.

Clarke EGA is a tool used to assess the clinical accuracy of blood glucose monitoring systems. It plots actual blood glucose measurements against predicted values on a scatter plot and divides the scatter plot into five regions: A, B, C, D, and E. Region A (clinical accuracy zone): Represents blood glucose values where the difference from the actual value

is $\leq 20\%$ or within the hypoglycemic range (< 70 mg/dL). Data falling within this zone can lead to correct clinical decisions. Zone B (clinically acceptable zone): Represents predicted values that deviate from actual values by $> 20\%$, but the resulting treatment is still clinically acceptable, requiring no intervention or benign intervention. Zones C, D, and E (clinically erroneous zones): Data in these zones may lead to unnecessary or dangerous treatment decisions, and clinical decisions based on these data may result in incorrect clinical treatment. Therefore, the higher the proportion of points falling within zones A and B, the more clinically reliable the predicted results are.

V. BASELINE AND EXPERIMENT SETTING

To validate the effectiveness of the proposed method, we set up a set of baseline models for the outlier detection and the blood glucose prediction tasks. For the outlier detection task they are: 1) GRU; 2) LSTM; 3) RNN; and 4) transformer. For the blood glucose prediction task they are: 1) LSTM; 2) dilated RNN (dilated RNN); 3) generative adversarial network (GAN); and 4) regression.

We implemented the proposed method and the baseline methods using the PyTorch deep learning framework in Python. The GRU has a hidden layer dimension of 256 and two layers. The LSTM and RNN have 128 hidden layer nodes. The transformer has four attention heads, six encoder and decoder layers, and a model dimension of 512. The dilated RNN has a hidden layer dimension of 32. The GAN generator consists of three layers of GRU, and the discriminator consists of three layers of 1D-CNN, with 32 hidden units per layer. In the regression model, the first layer is MLP/LSTM/PLSR, and the second layer is PLSR. In addition, we use the Adam optimizer, with dropout set to 0.1 and a learning rate of 0.0003. The operating system version used in the experiments was Windows 11, with an Intel 12600KF CPU, an NVIDIA RTX 4060Ti GPU, and 16 GB of VRAM.

In the experiments, we used CGM blood glucose values from the dataset as input information. During prediction, we utilized the historical data from the previous 2 h at the current time step to predict the blood glucose level at the corresponding time step 30-min later. When predicting the next time step, the historical window is also updated accordingly, and this process is repeated to complete the prediction of the entire time series.

A. Outlier Detection and Handling Results

Table I summarizes the successful detection rates S for the four injected fault types for ten of the Yuwell subjects. The performance of four configurations is compared: MT-TimesGlu (the full model in this article), without spatio-temporal attention (removing the attention module), without dual-task (removing joint learning), and baseline (the original single-task backbone model). The average detection rates of MT-TimesGlu for the four types of anomalies for all patients in the Yuyue dataset are 98.61%, 93.63%, 99.08%, and 92.26%, respectively. The results show that MT-TimesGlu is particularly effective in detecting short-term anomalies such as spikes

and steps, with an accuracy of 100% for many patients (e.g., 1935 030, 1935 090, 1935 290, etc.). This high sensitivity can be attributed to the model's multicycle spatiotemporal attention mechanism, which is able to effectively capture the mutation fluctuations and periodic local patterns of blood glucose. For long-term or subtle anomalies such as drift and PISA, the model still maintains robust performance, exceeding 95% for most patients, despite slightly lower detection rates. However, the drift detection rate for patient 1935 050 is relatively low, reflecting the challenge of modeling long-term dependencies and phase shifts across cycles. Overall, the MT-TimesGlu model showed strong generalization capabilities across different patients and abnormality types, achieving near-clinical-level detection accuracy for both transient and periodic abnormalities in CGM data.

The ablation results demonstrate that both the spatio-temporal attention mechanism and the dual-task learning framework play crucial yet complementary roles in achieving high anomaly detection accuracy. When the spatio-temporal attention module is removed (w/o spatio-temporal attention), the model shows a significant drop in performance for short-term anomalies such as spike and step, with detection rates decreasing from 99.16% to 92.06% and from 99.38% to 93.41%, respectively. This confirms that the multiscale attention design in MT-TimesGlu is vital for capturing local fluctuations and sudden glucose changes. On the other hand, removing the dual-task architecture (w/o dual-task) leads to a notable decline in the detection of long-term anomalies such as drift and PISA, where accuracy drops from 95.98% to 84.60% and from 96.99% to 89.45%, respectively. These results indicate that dual-task learning, by jointly optimizing prediction and anomaly detection, enhances the model's ability to identify subtle or prolonged abnormal patterns. Overall, MT-TimesGlu achieves superior performance by effectively integrating temporal attention for local feature enhancement and dual-task synergy for robust and generalizable representations across diverse fault types.

In addition, we also compared the performance of this method with GRU, LSTM, RNN, and transformer methods. The comparison results are shown in Table II. The MT-TimesGlu model achieved an average detection accuracy of 96.40%, significantly surpassing GRU (74.89%), LSTM (78.48%), RNN (83.39%), and transformer (80.25%), demonstrating outstanding generalization capability. RNN ranked second-best in performance due to its strong temporal modeling ability; however, its limited depth and absence of attention mechanisms restrict its adaptability to subtle anomaly patterns. GRU exhibited the weakest performance across all subjects. Although its gated architecture reduces model complexity, it struggles to differentiate fine-grained and complex outlier patterns in physiological data. LSTM showed moderate performance and performed well on certain subjects, but its redundant parameters may introduce noise in shorter sequences, thereby compromising robustness. The transformer, with its global self-attention mechanism, excels at capturing long-range dependencies but lacks sensitivity to local fluctuations, which limits its effectiveness in detecting short-term or abrupt CGM anomalies. In contrast, MT-TimesGlu integrates

TABLE I
ABLATION EXPERIMENTAL RESULTS OF OUTLIER DETECTION FOR FOUR TYPES OF FAULTS ON THE YUWELL DATASET

Algorithm	Patient (Yuwell)	Spike	Drift	Step	PISA
MT-TimesGlu	1935030	100%	98.11%	98.18%	92.86%
	1935050	98.04%	80.70%	98.08%	96.30%
	1935090	100.00%	96.43%	100.00%	100.00%
	1935190	98.08%	98.18%	100.00%	96.43%
	1935290	100.00%	100.00%	100.00%	100.00%
	1935310	100.00%	97.65%	100.00%	95.00%
	1935370	100.00%	98.18%	100.00%	96.15%
	1935450	98.18%	96.36%	100.00%	96.15%
	1935490	98.11%	98.18%	98.18%	100.00%
	Average	99.16%	95.98%	99.38%	96.99%
Baseline	1935030	98.15%	71.70%	96.36%	85.71%
	1935050	88.24%	50.88%	88.46%	88.89%
	1935090	94.34%	57.14%	94.44%	78.57%
	1935190	94.23%	61.82%	98.15%	96.43%
	1935290	100.00%	55.56%	94.55%	80.77%
	1935310	100.00%	73.17%	100.00%	95.00%
	1935370	91.07%	47.27%	88.68%	88.46%
	1935450	98.18%	72.73%	100.00%	84.62%
	1935490	98.11%	54.55%	96.36%	68.00%
	Average	95.80%	60.53%	95.22%	85.16%
w/o spatio-temporal attention	1935030	94.44%	84.91%	94.55%	78.57%
	1935050	84.31%	71.93%	90.38%	85.19%
	1935090	90.57%	76.79%	90.74%	75.00%
	1935190	88.46%	81.82%	90.74%	82.14%
	1935290	89.47%	68.52%	92.73%	80.77%
	1935310	97.67%	75.61%	92.68%	85.00%
	1935370	92.86%	78.18%	94.34%	88.46%
	1935450	94.55%	70.91%	100.00%	80.77%
	1935490	96.23%	74.55%	94.55%	80.00%
	Average	92.06%	75.91%	93.41%	81.77%
w/o dual-task	1935030	100.00%	90.57%	98.18%	89.29%
	1935050	94.12%	78.95%	96.15%	92.59%
	1935090	98.11%	83.93%	98.15%	82.14%
	1935190	98.08%	87.27%	100.00%	92.86%
	1935290	100.00%	72.22%	100.00%	88.46%
	1935310	100.00%	82.93%	100.00%	95.00%
	1935370	94.64%	83.64%	92.45%	92.31%
	1935450	100.00%	92.73%	100.00%	88.46%
	1935490	98.11%	89.09%	96.36%	84.00%
	Average	98.12%	84.60%	97.92%	89.45%

multiperiod spatio-temporal attention and dual-task optimization, enabling accurate detection of both transient and periodic irregularities, thereby delivering consistently superior performance across diverse subjects.

The experimental results of outlier imputation, evaluated using the PearsonCC and ICC, are presented in Supplementary Figs. S1 and S2. As shown in Fig. S1, MT-TimesGlu consistently achieves higher PearsonCC and ICC scores across all subjects, with average values exceeding 0.85, outperforming both the baseline and its ablated variants. Fig. S2 further demonstrates that, compared to other mainstream models, MT-TimesGlu maintains superior agreement between the imputed and true values, particularly under conditions of pronounced glucose fluctuations, indicating strong robustness and generalization capability. The experimental results of outlier imputation using PearsonCC and ICC are shown in supplementary Figs. S1 and S2. As shown in Fig. S1, MT-TimesGlu consistently achieves higher PearsonCC and

ICC scores across all subjects, with average values exceeding 0.85. Fig. S2 further demonstrates that, compared to other mainstream models, MT-TimesGlu maintains superior agreement between the imputed and true values, particularly under conditions of pronounced glucose fluctuations, indicating strong robustness and generalization capability.

B. Blood Glucose Prediction Results

In this study, blood glucose prediction experiments were conducted on both the Yuwell and OhioT1DM datasets. The evaluation metrics included RMSE, MAE, MARD, and Clarke EGA. Fig. 4 shows the prediction results of the proposed PGP-SARAN model for the next 15-, 30-, and 45 min of blood glucose concentration on the OhioT1DM and Yuwell datasets. Overall, the model can accurately track the changing trend of the real blood glucose concentration in most time periods, especially in the short-term prediction (15 min), showing extremely high consistency and fit, which verifies

TABLE II
EXPERIMENTAL RESULTS OF OUTLIER DETECTION USING DIFFERENT METHODS ON THE YUWELL DATASET

Patient (Yuwell)	MT-TimesGlu	GRU	LSTM	RNN	Transformer
1935010	96.39%	78.87%	88.66%	87.11%	72.16%
1935030	97.89%	72.11%	84.74%	75.79%	81.58%
1935050	92.51%	75.40%	74.33%	80.75%	68.45%
1935070	91.44%	76.47%	65.24%	80.75%	70.05%
1935090	98.95%	67.02%	83.25%	70.16%	77.49%
1935130	94.79%	70.31%	80.73%	82.81%	84.90%
1935150	95.34%	69.95%	72.54%	91.19%	68.39%
1935170	94.65%	57.75%	63.10%	79.14%	78.61%
1935190	98.41%	79.89%	67.72%	85.71%	84.13%
1935250	93.08%	49.23%	74.62%	85.38%	83.08%
1935270	96.34%	81.68%	85.34%	91.62%	86.39%
1935290	100.00%	84.38%	86.98%	87.50%	81.77%
1935310	98.62%	91.03%	91.03%	89.66%	93.10%
1935350	91.58%	61.05%	65.26%	71.05%	73.68%
1935370	98.95%	66.84%	78.95%	75.79%	67.37%
1935410	96.34%	93.90%	76.83%	89.02%	94.15%
1935430	98.44%	71.35%	79.17%	70.31%	85.94%
1935450	97.89%	93.68%	86.32%	93.16%	86.32%
1935470	97.94%	84.02%	86.08%	95.88%	86.60%
1935490	98.40%	72.87%	78.72%	85.11%	80.85%
Average	96.40%	74.89%	78.48%	83.39%	80.25%

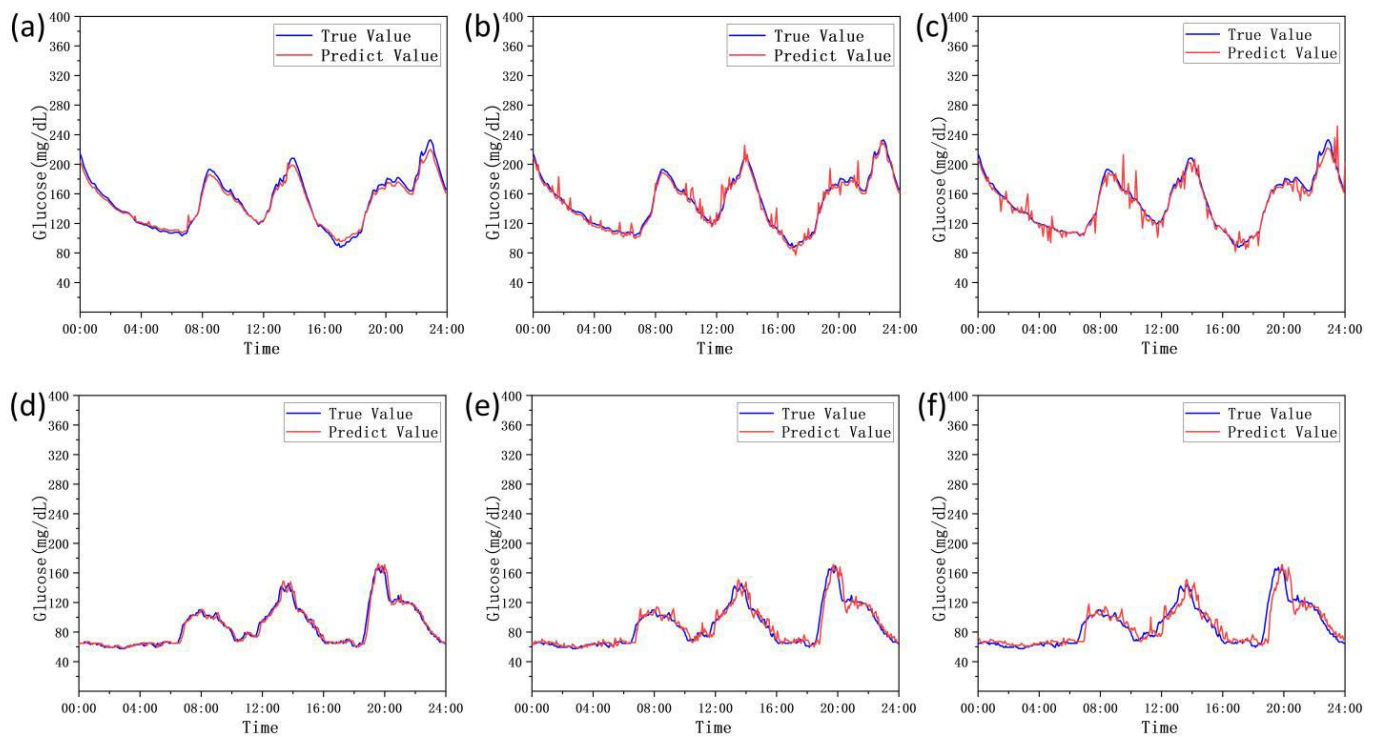


Fig. 4. Time plot of the blood glucose prediction performance of the PGP-SARAN model on a patient in both datasets. (a) 15-min prediction of Ohio. (b) 30-min prediction of Ohio. (c) 45-min prediction of Ohio. (d) 15-min prediction of Yuwell. (e) 30-min prediction of Yuwell. (f) 45-min prediction of Yuwell.

the effectiveness of its ability to model the dynamic changes of blood glucose. Overall, the model can accurately track the changing trend of the real blood glucose concentration in most time periods, especially in short-term prediction (15 min), showing extremely high consistency and fit, verifying the effectiveness of its ability to model dynamic changes in blood glucose. In the OhioT1DM dataset, the model successfully captured the main trend of blood glucose fluctuations

in the 15-min prediction, and the prediction curve almost coincided with the true value, indicating that the model has excellent short-term prediction capabilities. As the prediction time increases, although some periods (such as the afternoon trough or the sharp rise after a meal) have slight deviations or oscillations, the overall trend remains stable, and the red line can always return to the vicinity of the true track, showing good time robustness. Most of the areas where deviations

TABLE III
EXPERIMENTAL RESULTS OF RMSE, MAE, AND MARD FOR PATIENTS IN TWO DATASET

Patient ID	15min			30min			45min		
	RMSE	MAE	MARD	RMSE	MAE	MARD	RMSE	MAE	MARD
Yuwell (part)									
1935270	4.82	3.57	2.93%	8.37	6.35	5.21%	11.47	8.68	7.03%
1935370	4.29	2.99	3.30%	6.84	4.91	5.46%	9.75	6.95	7.78%
1935410	4.26	2.96	2.35%	7.93	5.47	4.29%	11.04	7.79	6.16%
1935430	5.45	4.04	3.56%	8.36	6.1	5.26%	11.29	8.37	7.22%
1935450	5.33	3.96	3.35%	8.23	6.11	5.18%	11.30	8.48	7.17%
1935470	3.93	2.72	2.35%	7.14	4.79	4.05%	10.09	6.48	5.34%
Average	6.27	4.20	3.76%	10.40	7.18	6.22%^	14.31	10.02	8.68%
Ohio									
540	10.92	7.86	5.45%	19.63	14.13	9.30%	31.76	23.77	16.59%
544	9.78	7.1	4.89%	14.98	10.19	6.17%	24.63	17.65	11.70%
552	9.19	6.66	5.11%	14.83	10.05	7.16%	23.88	17.76	13.92%
567	11.09	7.66	5.62%	19.04	12.35	7.99%	29.94	20.94	14.87%
584	13.26	9.17	6.20%	19.51	13.42	8.12%	28.68	20.84	13.52%
596	10.74	7.13	5.19%	14.89	9.94	6.75%	23.11	16.5	11.96%
559	10.41	6.85	4.33%	16.5	10.34	5.97%	25.83	18.18	11.98%
563	12.39	8.22	5.23%	16.44	11.09	6.19%	24.65	17.88	11.22%
570	8.92	6.18	3.18%	13.24	8.64	4.54%	21.13	15.03	7.86%
575	13.33	7.94	5.84%	19.64	11.98	8.39%	28.34	19.41	14.68%
588	10.11	7.27	4.24%	15.46	10.38	5.17%	23.97	17.2	10.00%
591	12.53	8.61	6.53%	18.76	13.07	9.82%	27.31	20.35	15.89%
Average	11.06	7.55	5.15%	16.91	11.30	7.13%	26.10	18.79	12.85%

occur are concentrated in the stage of rapid changes in blood glucose, which is due to the drastic influence of individual intervention factors such as diet, activity, and insulin. The model can still maintain reasonable accuracy in such cases, which has demonstrated strong generalization ability. In the Yuwell dataset, the model performs more stably, and the prediction curves under the three prediction times are highly coincident with the true value, even in the period of drastic blood glucose fluctuations (such as breakfast and after dinner). There is no significant deviation, indicating that the model has extremely strong prediction accuracy and robustness in this data environment. Considering that the dataset is collected under more standardized conditions and has a higher signal-to-noise ratio, PGP-SARAN can more fully exploit its temporal characteristics.

To comprehensively evaluate the forecasting performance of the proposed PGP-SARAN model across different prediction lengths and patient populations, we first examine its behavior over 15-, 30-, and 45-min horizons on two distinct datasets: Yuwell and OhioT1DM. As summarized in Table III, all three error metrics increase steadily with increasing prediction time, reflecting the expected increase in prediction difficulty. On the Yuwell dataset, the model maintains tight error control, with the average RMSE increasing only from 4.68 to 10.82 mg/dL and MARD increasing from 2.97% to 6.78%. In contrast, the OhioT1DM cohort has higher but still acceptable errors due to the lower sampling rate (5 min) and greater interpatient variability. In this dataset, patient 570 recorded the best results (RMSE = 13.24 mg/dL, MARD = 4.54%), while patient 540 had the largest error (RMSE = 19.63 mg/dL, MARD = 9.30%), a difference attributed to significant nonstationary glucose fluctuations that may be caused by undocumented lifestyle or insulin factors. Overall, the consistently low errors on the Yuwell dataset and clinically acceptable errors on the Ohio dataset suggest that PGP-SARAN has good gen-

eralization ability across different datasets and time frames, supporting its potential to achieve closed-loop glucose management in the real world.

To further visualize the consistency of the predictions, Fig. 5 shows the box plots of the three error metrics obtained using PGP-SARAN on the two datasets at 15-, 30-, and 45 min, respectively. For both cohorts the median error and interquartile range (IQR) grow monotonically with prediction lead time, confirming the anticipated loss of accuracy as the horizon lengthens. Yuwell box plots are consistently lower and tighter than Ohio box plots, indicating smaller central errors and lower dispersion. At 15 min, Yuwell median RMSE is below 6 mg/dL and IQR is below 1 mg/dL. Even at 45 min, Yuwell median MARD remains below 8%. Overall, the distributions corroborate the numeric averages in Table III, PGP-SARAN delivers tightly bounded, clinically acceptable errors on the in-house Yuwell data and retains manageable variability on the more challenging public OhioT1DM cohort. We also employ ClarkeEGA for error stratification in Supplementary Table S-III, which further demonstrates the clinical safety of the model. Table S-III shows the evaluation results of the PGP-SARAN method for the blood glucose prediction performance of patients in the Yuwell and OhioT1DM datasets under Clarke EGA when predicting for 30 min. Overall, the PGP-SARAN model achieved highly secure prediction performance in both datasets, with the vast majority of prediction points concentrated in A and B, which fully demonstrated the reliability and practical value of the model in actual deployment.

In addition, Fig. 6 shows the Clarke error grid analysis for two representative groups of subjects at 15-, 30-, and 45-min forecast periods. For the Ohio subject, 15- and 30-min forecasts place nearly all points within A, with a small spill-over into B, indicating clinically accurate or benign recommendations. Over the 45-min forecast period, there is a modest radial

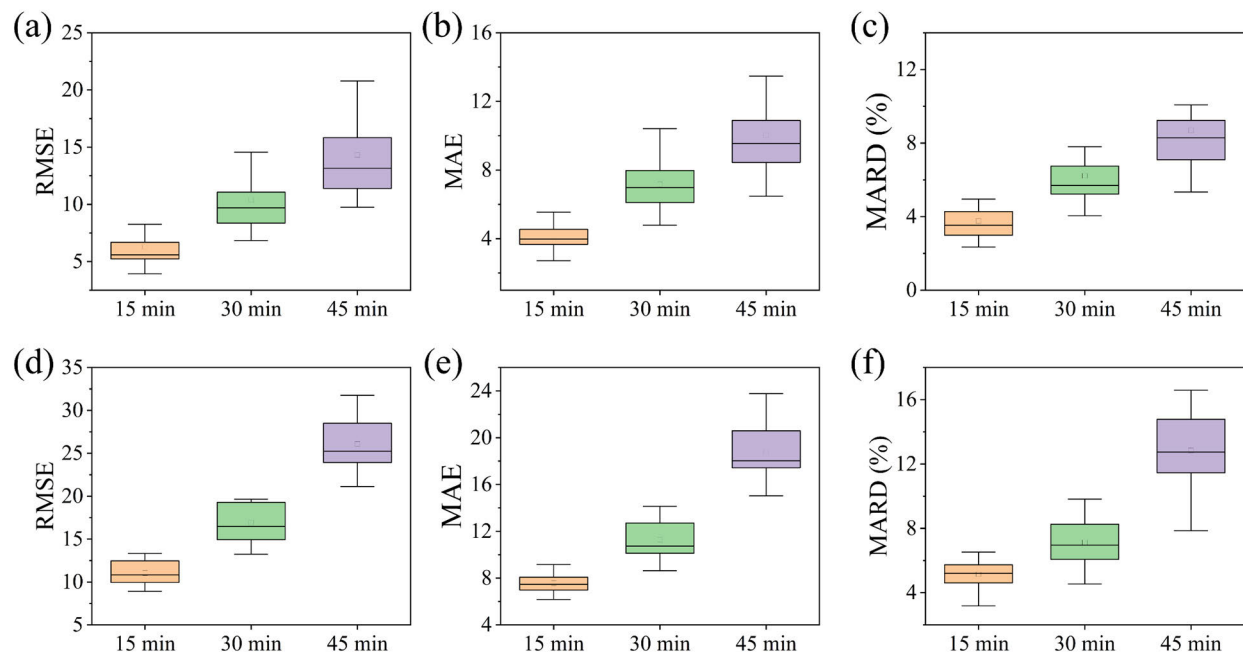


Fig. 5. Box plots of experimental results metrics for different datasets under different prediction ranges using PGP-SARAN. (a) RMSE for Yuwell dataset. (b) MAE for Yuwell dataset. (c) MARD for Yuwell dataset. (d) RMSE for OhioT1DM dataset. (e) MAE for OhioT1DM dataset. (f) MARD for OhioT1DM dataset.

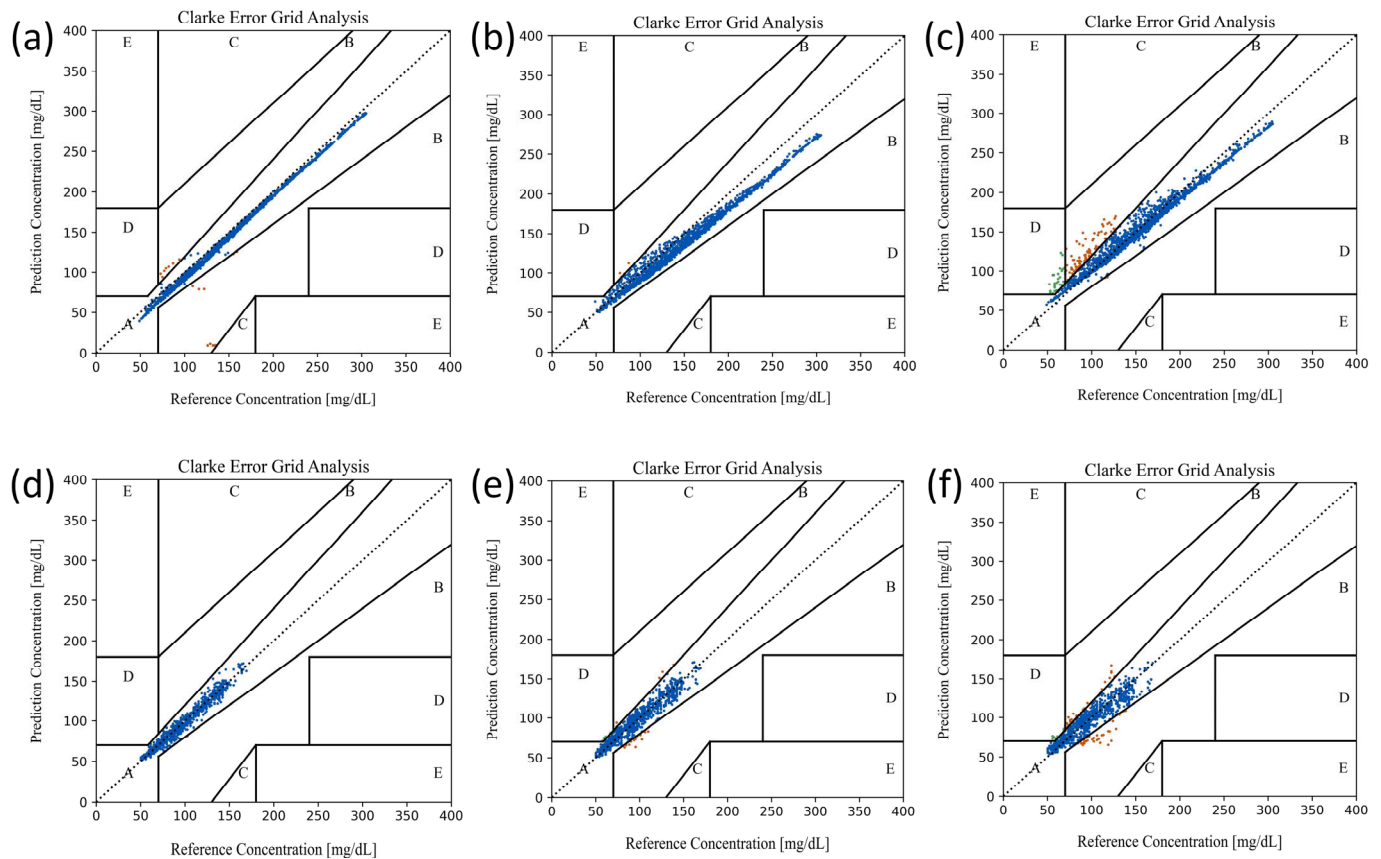


Fig. 6. Clarke EGA results of PGP-SARAN for blood glucose predictions in two datasets. (a) Ohio-15 min. (b) Ohio-30 min. (c) Ohio-45 min. (d) Yuwell-15 min. (e) Yuwell-30 min. (f) Yuwell-45 min.

spread with an increased proportion of points in B and some outliers moving into C and D, reflecting the greater uncertainty associated with a longer forecast period. The distribution of Yuwell subjects was tighter, with nearly all predictions in zone

A, a few in zone B, and no observations in risky D or E, across all ranges. These grids corroborate the numerical EGA statistics in Table IV, confirming that PGP-SARAN provides clinically safe guidance across short- and mid-term horizons

TABLE IV
EXPERIMENTAL RESULTS OF BLOOD GLUCOSE PREDICTION IN 30 MIN USING DIFFERENT METHODS

Patient ID	PGP-SARAN	LSTM ^[21]	Dilated RNN ^[30]	GAN ^[31]	Regression ^[32]
540	19.63	-	-	20.14	20.98
544	14.98	-	-	16.28	17.66
552	14.83	-	-	16.08	16.30
567	19.04	-	-	20.00	20.52
584	19.51	-	-	20.91	21.62
596	14.89	-	-	16.63	17.45
559	16.5	18.03	18.60	-	-
563	16.44	19.20	18.00	-	-
570	13.24	16.63	15.30	-	-
575	19.64	21.12	22.70	-	-
588	15.46	18.07	17.60	-	-
591	18.76	20.71	21.10	-	-

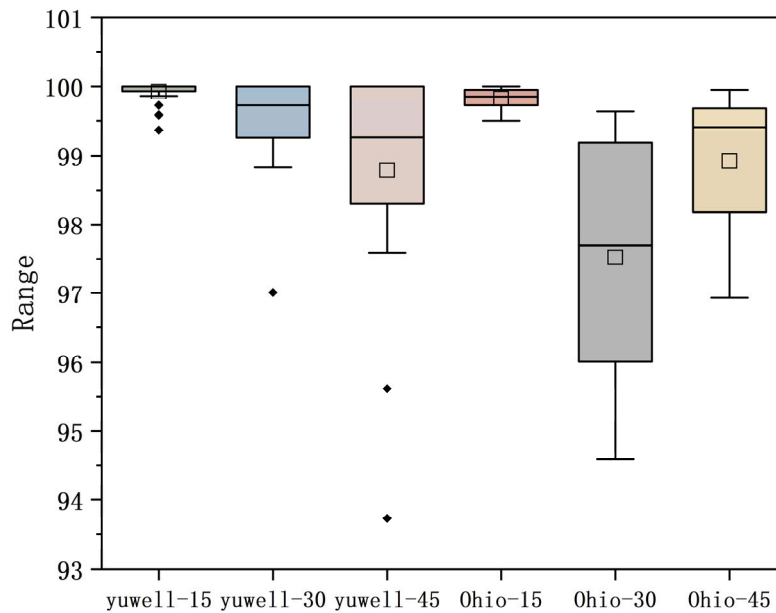


Fig. 7. Box plots of the proportion of ClarkeEGA scatter points falling into area A when the PGP-SARAN model predicts different lengths on two datasets.

and retains acceptable safety margins even under the more variable OhioT1DM conditions. Fig. 7 plots the percentage of each patient whose Clarke-EGA score falls in zone A. PGP-SARAN maintains almost maximal coverage of zone A on Yuwell and only slightly decreases on the more challenging OhioT1DM dataset, indicating that PGP-SARAN has strong decision-level safety even with increasing prediction length and data heterogeneity.

In the blood glucose prediction part, we compared the model in this article with the LSTM, dilated RNN, GAN, and regression model for 30-min prediction and used the RMSE as the evaluation indicator. The results are shown in Table IV. Since some studies only utilized one batch of the OhioT1DM dataset, this table conducts a cross-comparison based on the dataset batches. The PGP-SARAN model demonstrates a significant advantage over other models in experiments conducted on the OhioT1DM dataset. Compared to traditional LSTM and dilated RNN models, the self-attention mechanism of PGP-SARAN enhances modeling capabilities for long-range dependencies by dynamically allocating weights to prioritize nearby time points while capturing global dependencies. GAN improves model robustness through adversarial training, but its

unstable training process may lead to performance fluctuations. Regression models, lacking the ability to model nonlinear features, exhibit significant prediction errors in complex blood glucose fluctuations and have obvious limitations in capturing dynamic changes in blood glucose levels. In addition, PGP-SARAN converges quickly on a small amount of personal data through transfer learning strategies, reducing the need for large amounts of personal data. This strategy enables the model to better adapt to individual characteristics, thereby demonstrating high generalization ability across different patient data. Although GAN and LSTM perform well in some patients, they are highly dependent on data volume and have poor adaptability to extreme data points. PGP-SARAN significantly enhances the modeling capability of long-term dependencies through a global dynamic attention mechanism and a recurrent self-attention architecture and enhances individualized adaptability through a transfer learning strategy. It demonstrates higher robustness in predicting extreme data points and can effectively handle scenarios such as sudden drops or spikes in blood sugar levels. In contrast, LSTM and dilated RNN have limitations in modeling long-term dependencies and adapting to extreme data points, GAN still has room

for improvement in training stability, and regression models perform poorly due to their lack of nonlinear feature modeling capabilities. These analyses indicate that PGP-SARAN has significant performance advantages in blood glucose prediction tasks, particularly in short-term and medium-term prediction scenarios.

VI. CONCLUSION AND DISCUSSION

In this study, we proposed a two-stage framework for blood glucose analysis based on single-feature CGM data, integrating an outlier detection and correction module (MT-TimesGlu) with a recursive self-attention prediction model (PGP-SARAN) to ensure data integrity and enhance forecasting accuracy. Experimental results on the Yuwell dataset demonstrate that MT-TimesGlu effectively detects various types of anomalies—spike, drift, step, and PISA—with an average detection accuracy exceeding 95%. The corrected values maintain high consistency with ground truth, reflected by PearsonCCs and ICC values consistently above 0.85, thereby reducing the negative impact of anomalies on downstream modeling. Building on this foundation, the PGP-SARAN model leverages a global dynamic attention mechanism and a recursive self-attention architecture to accurately capture long-term dependencies. It further improves adaptability to individual patient characteristics through a transfer learning strategy, significantly outperforming baseline methods including LSTM, dilated RNN, GAN, and regression. The proposed approach achieves consistently low RMSE values and high proportions of Clarke EGA region A predictions on both the Yuwell and OhioT1DM datasets, validating its robustness and clinical reliability in short- and medium-term glucose prediction tasks. Future work will focus on further optimizing and refining the proposed model by exploring alternative architectural variants and hyperparameter configurations to improve the accuracy of long-term blood glucose trend prediction. In addition, we aim to develop models capable of generating uncertainty-aware outputs, thereby enhancing clinical interpretability and supporting safer decision-making in real-world applications.

REFERENCES

- [1] M. Fralick, A. J. Jenkins, K. Khunti, J. C. Mbanya, V. Mohan, and M. I. Schmidt, "Global accessibility of therapeutics for diabetes mellitus," *Nature Rev. Endocrinol.*, vol. 18, no. 4, pp. 199–204, Apr. 2022.
- [2] B. Zhou et al., "Worldwide trends in diabetes prevalence and treatment from 1990 to 2022: A pooled analysis of 1108 population-representative studies with 141 million participants," *Lancet*, vol. 404, no. 10467, pp. 2077–2093, 2024.
- [3] R. Kasole et al., "Traditional medicine and its role in the management of diabetes mellitus: Patients and herbalists perspective," *Evidence-Based Complementary Alternative*, vol. 2019, no. 1, 2019, Art. no. 2835691.
- [4] D. Olczuk and R. Priefer, "A history of continuous glucose monitors (CGMs) in self-monitoring of diabetes mellitus," *Diabetes Metabolic Syndrome, Clin. Res. Rev.*, vol. 12, no. 2, pp. 181–187, Apr. 2018.
- [5] Z. Pu et al., "A thermal activated and differential self-calibrated flexible epidermal biomicrofluidic device for wearable accurate blood glucose monitoring," *Sci. Adv.*, vol. 7, no. 5, Jan. 2021, Art. no. eabd0199.
- [6] P. D. G. Miles, M. Levisetti, D. Reichart, M. Khoursheed, A. R. Moossa, and J. M. Olefsky, "Kinetics of insulin action in vivo: Identification of rate-limiting steps," *Diabetes*, vol. 44, no. 8, pp. 947–953, Aug. 1995.
- [7] X. Yu, X. Sun, Y. Zhao, J. Liu, and H. Li, "Fault detection of continuous glucose measurements based on modified k-medoids clustering algorithm," *Neural Comput. Appl.*, pp. 1–10, Oct. 2020.
- [8] K. Turksoy, A. Roy, and A. Cinar, "Real-time model-based fault detection of continuous glucose sensor measurements," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1437–1445, Jul. 2017.
- [9] X. Yu et al., "Fault detection in continuous glucose monitoring sensors for artificial pancreas systems," *IFAC-PapersOnLine*, vol. 51, no. 18, pp. 714–719, 2018.
- [10] H. Wang, M. J. Bah, and M. Hammad, "Progress in outlier detection techniques: A survey," *IEEE Access*, vol. 7, pp. 107964–108000, 2019.
- [11] Y. Zhao, H. Li, X. Yu, N. Ma, T. Yang, and J. Zhou, "An independent central point OPTICS clustering algorithm for semi-supervised outlier detection of continuous glucose measurements," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103196.
- [12] J. Li, I. Tobore, Y. Liu, A. Kandwal, L. Wang, and Z. Nie, "Non-invasive monitoring of three glucose ranges based on ECG by using DBSCAN-CNN," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 9, pp. 3340–3350, Sep. 2021.
- [13] E. Idi, E. Manzoni, A. Facchinetti, G. Sparacino, and S. D. Favero, "Unsupervised retrospective detection of pressure induced failures in continuous glucose monitoring sensors for T1D management," *IEEE J. Biomed. Health Informat.*, vol. 29, no. 2, pp. 1383–1396, Feb. 2025.
- [14] R. N. Bergman et al., "Quantitative estimation of insulin sensitivity," *Amer. J. Physiol.*, vol. 236, no. 6, pp. E667–E677, 1979.
- [15] J. T. Sorensen, "A physiologic model of glucose metabolism in man and its use to design and assess improved insulin therapies for diabetes," Ph.D. dissertation, Dept. Chem. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 1985.
- [16] E. I. Georga et al., "Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 1, pp. 71–81, Jan. 2013.
- [17] E. I. Georga, J. C. Príncipe, and D. I. Fotiadis, "Short-term prediction of glucose in type 1 diabetes using kernel adaptive filters," *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 27–46, Jan. 2019.
- [18] Y. Lu, A. V. Gribok, W. Kenneth Ward, and J. Reifman, "The importance of different frequency bands in predicting subcutaneous glucose concentration in type 1 diabetic patients," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1839–1846, Aug. 2010.
- [19] T. Hamdi, J. Ben Ali, V. Di Costanzo, F. Fnaiech, E. Moreau, and J.-M. Ginoux, "Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm," *Bio-cybernetics Biomed. Eng.*, vol. 38, no. 2, pp. 362–372, 2018.
- [20] K. Li, J. Daniels, C. Liu, P. Herrero, and P. Georgiou, "Convolutional recurrent neural networks for glucose prediction," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 603–613, Feb. 2020.
- [21] M. F. Rabbay, Y. Tu, M. I. Hossen, I. Lee, A. S. Maida, and X. Hei, "Stacked LSTM based deep recurrent neural network with Kalman smoothing for blood glucose prediction," *BMC Med. Informat. Decis. Making*, vol. 21, no. 1, p. 101, Dec. 2021.
- [22] P. Madan et al., "An optimization-based diabetes prediction model using CNN and bi-directional LSTM in real-time environment," *Appl. Sci.*, vol. 12, no. 8, p. 3989, Apr. 2022.
- [23] A. B. Sharma, L. Golubchik, and R. Govindan, "Sensor faults: Detection methods and prevalence in real-world datasets," *ACM Trans. Sensor Netw.*, vol. 6, no. 3, pp. 1–39, Jun. 2010.
- [24] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "TimesNet: Temporal 2D-variation modeling for general time series analysis," 2022, *arXiv:2210.02186*.
- [25] A. Weissman and O. Binah, "The fractal nature of blood glucose fluctuations," *J. Diabetes its Complications*, vol. 28, no. 5, pp. 646–651, Sep. 2014.
- [26] J. Yang et al., "Analysis of blood glucose fluctuation and its influencing factors in patients with type 2 diabetes," *Chin. J. Practical Nursing*, vol. 2018, pp. 401–406, Apr. 2018.
- [27] C. Gao, N. Zhang, Y. Li, F. Bian, and H. Wan, "Self-attention-based time-variant neural networks for multi-step time series forecasting," *Neural Comput. Appl.*, vol. 34, no. 11, pp. 8737–8754, Jun. 2022.
- [28] M. Sinha et al., "A comparison of time delay in three continuous glucose monitors for adolescents and adults," *J. Diabetes Sci. Technol.*, vol. 11, no. 6, pp. 1132–1137, Nov. 2017.
- [29] C. Marling and R. Bunescu, "The OhioT1DM dataset for blood glucose level prediction: Update 2020," in *Proc. CEUR Workshop*, vol. 2675, 2020, pp. 4–71.
- [30] T. Zhu, K. Li, J. Chen, P. Herrero, and P. Georgiou, "Dilated recurrent neural networks for glucose forecasting in type 1 diabetes," *J. Healthcare Informat. Res.*, vol. 4, no. 3, pp. 308–324, Sep. 2020.
- [31] T. Zhu et al., "Blood glucose prediction for type 1 diabetes using generative adversarial networks," in *Proc. CEUR Workshop*, vol. 2675, 2020, pp. 90–94.
- [32] H. Nemat et al., "Data fusion of activity and CGM for predicting blood glucose levels," in *Proc. Knowl. Discovery Healthcare Data, CEUR Workshop*, vol. 2675, 2020, pp. 120–124.