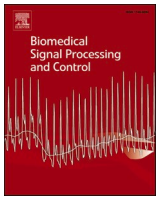
Biomedical Signal Processing and Control 82 (2023) 104552

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Short-term prediction method of blood glucose based on temporal multi-head attention mechanism for diabetic patients☆

Guanci Yang [a,b], Saisai Liu [a], Yang Li [a,*], Ling He [a]

[a] Key Laboratory of Advanced Manufacturing Technology of the Ministry of Education, Guizhou University, Guiyang 550025, China
[b] State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China

ARTICLE INFO

Keywords:
Blood glucose prediction
Multi-head attention mechanism
Spatiotemporal features
Sequence data

ABSTRACT

The hyperglycemic state of people with diabetes can lead to metabolic and healthy disturbances in the body. Diabetes is mainly treated clinically by conservative treatment, which requires frequent and continuous measurement of blood glucose concentration. Accurate blood glucose prediction plays an important role in the future blood glucose management of patients. To improve the accuracy of blood glucose prediction, this paper proposes a short-term prediction method of blood glucose based on temporal multi-head attention mechanism for diabetic patients (PBGTAM). Firstly, a detection algorithm of abnormal blood glucose based on Adaptive Density-Based Spatial Clustering of Applications with Noise is proposed by using an autonomous neighborhood parameter selection method. Secondly, the imputation algorithm based on feature engineering is proposed to fill the missing blood glucose values. Thirdly, we propose temporal multi-head attention model to obtain global and local spatiotemporal features from sequence data, in which a temporal series features module is designed to insight the detail feature so as to keep the useful information contained in the details, and an intensity correlation module is used to obtain global features of sequence data and then to achieve strong correlation between blood glucose data, and a sequence features module is employed to learn temporal sequence features between sequence data to guarantee continuous prediction of future blood glucose values. Finally, the comparative results show that PBGTAM outperforms six state-of-the-art methods in terms of the overall prediction root mean square error, forecasting accuracy for warning, and clinical consistency with 20.57, 84.35% and 85.18%, respectively.

1. Introduction

Diabetes is a common metabolic disease. Long-term high blood sugar levels in diabetic patients will lead to metabolic disorders, and easily lead to a series of complications, which will seriously affect the patient's health. According to World Health Organization statistics, in 2019, about 1.5 million people worldwide died because of diabetes, and more than 450 million people are living with diabetes. Projections also show that the number of diabetics worldwide will rise to about 700 million by 2045, indicating that diabetes has become one of the primary causes of deaths worldwide [1]. Based on current medical technology, diabetes is still incurable, and diabetic patients can only control their blood glucose levels by long-term medication and insulin injection [2]. Usually, patients need regular and frequent blood glucose detection. The methods of blood glucose detection are mainly divided into venous blood glucose detection and rapid blood glucose meter detection. Both need to obtain blood samples through needle puncture, which brings pain to patients. In addition, as a population with a high incidence of diabetes, middle-aged and elderly patients have trouble going to the hospital frequently due to their mobility, which will bring trouble to them and their children. Therefore, accurate blood glucose prediction can bring great convenience. It can predict the blood glucose level of patients in real time, effectively reduce the number of blood glucose detection, relieve patients' pain, and play an important role in the future blood glucose management and continuous treatment of patients.

To reduce the number of blood samples and master the changes of blood glucose at the same time, some researchers studied the blood glucose prediction method based on the physiological model [3].

☆ This work is supported in part by the National Natural Science Foundation of China under Grant Nos. 62163007, and the Science and Technology Foundation of Guizhou Province under grant [2020]4Y056, PTRC [2020]6007, and [2021]439, and Graduate Research Innovation Fund of Guizhou Province under Grant YJSCXJH [2020]090.

* Corresponding author.
E-mail addresses: gcyang@gzu.edu.cn (G. Yang), liyanggzu@163.com (Y. Li).

https://doi.org/10.1016/j.bspc.2022.104552
Received 24 September 2022; Received in revised form 14 December 2022; Accepted 26 December 2022
Available online 5 January 2023

Considering factors such as glucose metabolism, insulin and carbohydrate absorption, the parameters of the blood glucose prediction model require prior knowledge to be adjusted. Later, deep learning brought state-of-the-art performance to task in blood glucose prediction field. By using machine learning, Nie et al. [4] proposed a blood glucose prediction method, which is in better agreement with the reference glucose monitor and has more application prospects. It was also used to build a blood glucose prediction model for hyperglycemia and hypoglycemia [5]. Prediction models based on Long short-term memory (LSTM) [6–8] can learn long-term dependencies, allowing neural networks to combine their memory and input to improve the accuracy of predictions, the convolutional neural network (CNN) [9] model has better prediction accuracy and robustness. However, these prediction methods usually focus only on the patient's glycemic history and ignore insulin and carbohydrate intake factors. Some new models significantly improve the performance of blood glucose prediction [10]. The temporal convolutional network (TCN) [11] has feature extraction and time series modeling ability, but their subsequent output can only be predicted based on the previous input sequence information, and the sequence values that has not been input cannot be used. The prediction method based on bi-directional long short-term memory (BiLSTM) [12] can make full use of backward feature information from back to front, but it is not strong in data detail insight and easily loses useful information. The multi-head attention mechanism [13] can capture details.

Abnormal data collected from blood glucose can affect the accuracy of predictions. Abnormal data can be caused by a reaction to a foreign body in the tissue fluid, a malfunctioning blood glucose meter (CGM), or loss of collected data. To mitigate prediction errors caused by data abnormalities, Jia et al have focusing on data preprocessing and investigated detection models that attenuate the influence of outliers [14]. However, this method makes it difficult to describe uncertain physiological changes and random noise, leading to poor generalizability of the detection model.

The data-driven detection method [15] detects abnormal data through threshold and model training required labeled samples, but the reality is that it is often difficult to provide a large number of labeled samples. Methods based on unsupervised learning [16–17] do not need any data labeling. The core of unsupervised learning is clustering, which can model the data without any training samples. The clustering algorithm mainly divides the data set into multiple classes composed of similar objects by calculating the distance between samples and groups. It is competitive in the detection of outliers, but less stable.

To improve the accuracy of blood glucose prediction, this paper proposes a short-term prediction method of blood glucose based on temporal multi-head attention mechanism for diabetic patients (PBGTAM). The main contributions are summarized as follows.

- We proposed the short-term prediction method of blood glucose based on temporal multi-head attention mechanism for diabetic patients (PBGTAM) to improve the accuracy of blood glucose prediction. The PBGTAM is characterized by the ability of obtaining global and local spatiotemporal features from sequence data.
- We designed an unsupervised detection algorithm of abnormal blood glucose based on ADBSCAN without prior knowledge to accurately eliminate outliers, which independently updates the neighborhood parameter according to the characteristics of the data interval.
- Proposed imputation algorithm based on feature engineering for missing values to fill the missing blood glucose values.
- We compared PBGTAM with six outstanding methods, and the results show that PBGTAM outperforms these state-of-the-art methods in terms of generalizability, robustness, prediction accuracy, and clinical consistency.

The remainder is arranged as follows. Section 2 introduces the related work, Section 3 details temporal multi-head attention model PBGTAM, and Section 4 focuses on experimental results and analysis. Section 5 concludes.

2. Related work

Current research on the use of artificial intelligence to achieve blood glucose control mainly focuses on two aspects: blood glucose data preprocessing and prediction model design. From the perspective of feature engineering, high-quality data needs to eliminate discrete outliers in the original sequence data to ensure the stability of algorithm performance [18]. Detection based on the causes of outlier generation often takes into dynamic changes in account blood glucose (BG) and interstitial glucose (IG) concentration [19–20], calibration errors [21], and random noise [22], and the model constructed is mostly combinatorial. In particular, considering the human physiological changes and randomness of the environment, the model has poor generalizability and is only suitable for special sensors. Different from the detection based on the cause of the abnormal values, the data-driven error detection only needs historical data to estimate the difference between the measured values and the true values and then judge whether the threshold values are exceeded to complete fault detection.

In early studies, the wavelet [23] was used for data anomaly detection, but the detection effect of wavelet methods is limited by the degree of error. The blood glucose values predicted by Kalman filtering [24–25] were compared with the sampled values to make judgments, but the settings of Kalman parameters and the corresponding judgment criteria can affect the final detection results. Cinar et al. [26] used principal component analysis for outlier detection. However, when multicollinearity exists among original variables, it performed poorly because it cannot effectively remove information overlap. The autoregressive moving-average model [15] was also used to estimate the blood glucose values the next time and detect the fault through the error threshold. However, such data-driven anomaly detection heavily depends on a large number of historical data, and its prediction accuracy is very sensitive to the size and distribution of the dataset.

By contrast, as an unsupervised learning method, clustering can reveal the inherent properties and laws of the data without labeling the samples. Yu et al. [16] used a k-medoids clustering algorithm to detect outliers. The detection effect is significantly competitive, but this type of algorithm needs to randomly select k data samples as center points, which causes the clustering results to fluctuate greatly. Lim et al. [27] proposed density-based applied spatial clustering with noise (DBSCAN). DBSCAN clustering algorithm has more stable results in detecting and identifying outliers and is widely used in various fields [17], but it requires prior knowledge to set neighborhood parameters (ε, MinPts). Therefore, it is also worth investigating how to dynamically set ε with as few samples as possible in a large number of labeled samples. These studies mentioned above are summarized in Table 1.

In terms of blood glucose prediction, Eren-Oruklu et al. [28] proposed an adaptive univariate model based on the ARMA model for blood glucose prediction, which uses a single historical data to predict the current state, but the model was not stable. Therefore, using multivariable to predict can reduce the dependence on a single data [29]. Georga et al. [30] proposed a recursive multivariable kernel adaptive filtering method for personalized blood glucose short-term prediction. Wang et al. [31] believed that insulin dose and dietary intake have an impact

Table 1
Classification and comparison of outliers detection algorithm.

Methods	Advantage	Disadvantage
detection based on the causes of outlier generation [19–22]	consider the dynamic changes of various influencing factors	poor generalizability
data-driven error detection [15,23–26]	simple calculation	sensitive to the size and distribution of dataset
unsupervised learning method [16–17,27]	without labeling the samples	result fluctuates greatly

on blood glucose fluctuations and incorporated daily events, such as glucose, insulin [32], diet, and exercise [33], into the model to make it more effective. Although this type of model improves the prediction accuracy because the dependence on a single variable is reduced, the generalizability is poor. Practice has proved that [34] using only one specific method to predict blood glucose can produce one-sided results, and combining various prediction methods can give full play to their respective advantages and optimize prediction results.

Rabby et al. [24] used a deep recurrent neural network model with a stacked LSTM for glucose prediction, which expanded the length of the input data compared to a single LSTM. However, the LSTM can only capture the features of the data from front to back and not from back to front while the BiLSTM can bridge this drawback. Thus, another study [12] used BiLSTM layers to predict glucose levels in different ranges of blood with higher prediction accuracy compared to the LSTM. To expand the depth of the CNN input data, some studies [9,35–36] proposed a deep DCNN that handled multi-dimensional long signals through expanded convolution and performed better than shallow networks. Xie et al. [25] used TCN to predict the BG level and evaluated the prediction performance of LSTM, ARX, and SVR models. TCN was found to be more robust in the BG trajectory. Li et al. [37] creatively proposed the combination of CNN and LSTM, which has amazing performance in prediction accuracy. Table 2 shows these prediction algorithms mentioned above.

In summary, in terms of blood glucose data preprocessing, abnormal detection of the abovementioned supervised learning method relies on labeled data. However, clustering algorithms that do not require labels often require higher prior knowledge. On the whole, the size and distribution of the dataset directly affect the prediction accuracy of the above two types of methods. The ADBSCAN clustering algorithm we improved does not require a large amount of historical data in outliers detection. And compared with the DBSCAN, neighborhood parameters can update autonomously. At the same time, in terms of blood glucose prediction, most of the existing work only considers a single factor, and its prediction accuracy and generalizability are difficult to guarantee. Meanwhile, the methods based on TCN or LSTM cannot obtain the global and timing characteristics of sequence data at the same time, leading to poor performance in long-term blood glucose prediction. PBGTAM keeps the useful information contained in the details by obtaining global and local spatiotemporal features from sequence data.

Table 2
Classification and comparison of blood glucose prediction algorithm.

Methods	Advantage	Disadvantage
use single historical data [28]	simple calculation	not stable and produce one-sided results
use multivariable historical data [29–30]	optimize the prediction results	poor generalizability
LSTM [6–8]	can learn long-term dependencies	ignore insulin and carbohydrate intake factors
stacked LSTM [24]	expanded the length of the input data	capture the features of the data in one direction
BiLSTM [12]	capture the features of data in bi-direction	not strong in data detail insight and easily loses useful information
TCN [25]	great feature extraction and time series modeling ability	cannot use the sequence values that has not been input
CNN [9]	better prediction accuracy and robustness	focus only on the patient's history data
deep DCNN [9,35–36]	performed better than shallow networks	more computational cost
the combination of CNN and LSTM [37]	amazing performance in prediction accuracy	performance degrades as the prediction horizon increases

3. Temporal multi-head attention model PBGTAM

3.1. Proposed architecture of PBGTAM

For blood glucose prediction, to gain insight into the details of the data and obtain the global and timing characteristics of the sequence data, this section proposes the architecture of short-term prediction of blood glucose based on temporal multi-head attention mechanism for diabetic patients (PBGTAM) (Fig. 1). First PBGTAM receives data from BG (blood glucose), Basal, Bolus and Carbs(carbohydrates). Among them, in order to meet the metabolism needs of diabetes patients in the fasting state, the preset continuous infusion of short-term insulin is called Basal, while Bolus is used to provide the insulin needed for carbohydrate consumption. The data is then processed, mainly to remove outliers caused by noise pollution and to fill in missing values. Finally, global and local spatiotemporal features are extracted.

3.2. Detection algorithm of abnormal blood glucose based on ADBSCAN

Eliminating outliers in blood glucose monitoring data ensures high-quality blood glucose predictions. *K*-means is a common density-based unsupervised machine learning clustering algorithm, but its application scope and performance are limited [38]. DBSCAN can find clusters of arbitrary shape in the presence of noise points. Therefore, this study refers to DBSCAN for abnormal blood glucose detection.

The effectiveness of the DBSCAN clustering algorithm depends on the neighborhood parameters $(\varepsilon, MinPts)$, which depend heavily on prior knowledge. In fact, due to the difference in body constitution, the changes in blood sugar are also different. In this regard, this study designs the detection algorithm of abnormal blood glucose based on adaptive DBSCAN, which autonomously selects the neighborhood parameters ε according to the characteristics of the data interval. Algorithm 1 is the pseudocode.

Algorithm 1 Detection algorithm of abnormal blood glucose based on ADBSCAN

Input: dataset $G = \{g_1, g_2, g_3, \cdots, g_m\}$; the number of intervals w
Output: Outlier $G_{out} = \{g_1, \cdots, g_i\}$
Initialize $N = \varnothing$, $M = \varnothing$
For $i = 1$ *to* m
  Initialize $D_i = \varnothing$
  For $j = 1$ *to* m
    $d(x_j, x_i) = \|x_j - x_i\|_2$, calculate the Euclidean distance $d(x_j, x_i)$;
    $D_i = D_i \bigcup d$;
  $d_{zone} = \frac{\max(D_i) - \min(D_i)}{w}$, divide D_i into w sets, $D_i^f \in [\min(D_i) + f^* d_{zone}, \min(D_i) + (f+1)^* d_{zone}], f \in (0, w-1)$
  $n_i^f = \left|D_i^f\right|, m_i^f = mean\left(D_i^f\right)$, calculate the mean distance m_i^1
  if $n_i^1 > l/w$
  $N = N \bigcup n_i^1$, $M = M \bigcup m_i^1$, count the element g_i that satisfies the condition
$\varepsilon = mean(M), MinPts = 3$;
Initialize the core object collection: $O = \varnothing$
For $i = 1$ *to* m
  $P_\varepsilon(x_i) = \{x_i \in X | d(x_j, x_i) \le \varepsilon\}$, determine the neighborhood $P_\varepsilon(g_i)$ of g_i;
  if $|P_\varepsilon(g_i)| > MinPts$, then
    Add g_i to the core object collection $O : O = O \cup g_i$
Initialize the number of clusters: k = 0,
While $O \neq \varnothing$
  Initialize $P_k = \varnothing$
  Take any element $o \in O, P_k = P_k \cup o$
  if $|P_\varepsilon(o)| \rangle MinPts$, then $P_k = P_k \cup P_\varepsilon(o)$, and remove duplicate samples, $B = P_\varepsilon(o) \bigcap O$
    if $B \neq \varnothing$, $O = O - B$, take elements in B and repeat Step 18
    else
      $k = k + 1$, generate clusters $C_k = P_k$, go back to Step 17
Output $G_{Out} = G - C$

In Algorithm 1, for the blood glucose dataset $G = \{g_1, g_2, g_3, \cdots, g_m\}$, Step 4 calculates the distance d_{ij} between the element g_i and other elements in the G to obtain the distance set D_i. In Step 6, we divide D_i into w sets $\{D_i^1, D_i^2, \cdots, D_i^w\}$ according to the difference between the maximum

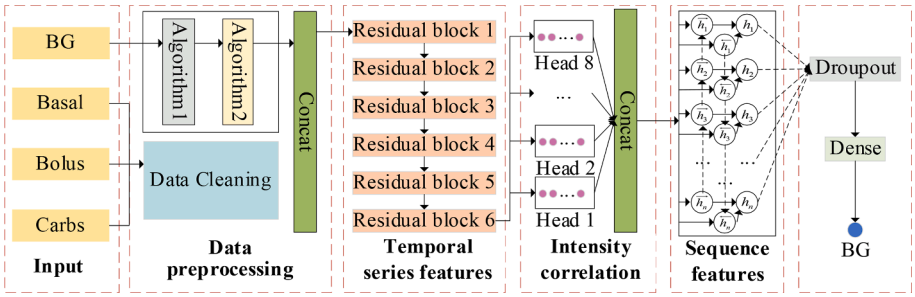


Fig. 1. Architecture of PBGTAM.

and minimum distances and calculate the mean distance m_i^1 based on the number of elements n_i^1 in the minimum distance set D_i^1. Then, the other elements in the G set are calculated according to the same method. In Step 8, we count the element g_i that satisfies the condition $n_i^1 > \frac{l}{w}$ (where l is the number of samples in the G, w is the number of subsets, and $w = floor(\frac{l}{MinPts})$). After this, we can get the $N = \{n_1^1, n_2^1, \cdots, n_q^1\}$ and $M = \left\{m_1^1, m_2^1, \cdots, m_q^1\right\}$ (where q is the number of elements satisfying the condition) and then calculate the mean of M as this sample neighborhood parameter ε. ADBSCAN selects the neighborhood parameters ε in this way. Since the patient's blood glucose does not change much in 20 min, we set the neighborhood density threshold $MinPts = 4$ in Step 8 and output the outlier set according to Steps 9–22.

The basic time complexity of DBSCAN is O (m* the time it takes to find a point in the ε field), where m is the number of points. In the worst case, the time complexity is $O(m^2)$. Our proposed ADBSCAN selects the neighborhood parameter ε autonomously by calculating the mean value of M as the neighborhood parameter ε of the sample. This improved operation results in the time complexity consumption of $O(m)$, which is almost negligible. Therefore, the worst time complexity of ADBSCAN is $O(m^2 + m)$.

3.3. Imputation algorithm based on feature engineering for missing values

For the problem of missing blood glucose values, inspired by feature engineering, we proposed the imputation algorithm based on feature engineering for missing value (as shown in Algorithm 2). In Algorithm 2, in steps 1–5, we mark the missing values' position rank according to the continuous missing values. Steps 6–11 are the forward estimation process of the missing values, and steps 12–9 are the backward estimation process of the missing value. In Step 7, for missing values of lower rank, we calculate the current value using g_{i-1} and g_{i-2} before the timestamp. To reduce the accumulation of errors, for missing values of higher rank, the same day blood glucose data $G_{today} = \{g_{i-index}, g_{i-index+1}, \cdots, g_{i-1}\}$ are introduced as important information to fill g_i besides g_{i-1} and g_{i-2}.

Algorithm 2. Imputation algorithm based on feature engineering for missing values

Input: CGM data with partial missing values, $G = \{g_1, g_2, g_5, \cdots, g_n\}$
Output: CGM data with no missing values, $G_* = \{g_1, g_2, g_3, \cdots, g_n\}$

1 for $i = 1$ to n
2 if $g_i = \varnothing$, then $g_i = 1$, $rank_i = t, t = t + 1$
3 else $t = 0$
4 for $i = 1$ to n
5 if $g_i = 1$
6 if $rank_i < 4$, then $g_i = 2g_{i-1} - g_{i-2}$;
7 else

$$index = i - floor\left(\frac{i}{288}\right) * 288, s = q * 288 + index,$$

$$q \in \left(0, floor\left(\frac{i}{288}\right)\right); g_1^* = 2g_{i-1} - g_{i-2};$$

$$g_2^* = mean(sum(g_j)), j \in (i - index, i - 1);$$

$$g_3^* = mean\left(sum(g_{(s)})\right);$$

$$g_i = g_1^* * 0.9^{rank_i} + 0.5 * (g_2^* + g_3^*)(1 - 0.9^{rank_i});$$

(continued on next column)

(continued)

8 if $g_i = 0, g_{i-1} = 1$, then $k = floor(0.5 * rank_i)$
9 for $j = 1$ to k
10 if $rank_{i-j} < 4$, then $g_{i-j} = 2 * g_{i-j+1} - g_{i-j+2}$
11 else $g_1^* = 2 * g_{i-j+1} - g_{i-j+2}$, g_2^* and g_3^* are the same for Step 9
$g_{i-j} = g_1^* * 0.9^{rank_{i-j}} + 0.5 * (g_2^* + g_3^*)(1 - 0.9^{rank_{i-j}})$
12 $OutputG_*$

In addition, due to the laws of human living habits, the blood sugar concentration also conforms to the regular changes. Therefore, based on the periodic and progressive changes in blood glucose levels, we make the following assumption that the blood glucose value at time t behaves similarly to the behavior at time t in the case of similar living patterns in adjacent n days, the blood glucose $G_{st} = \left\{g_t, g_{t+288}, \cdots, g_{t+288*q}\right\}$ at the same time in the historical data is also taken into account. Since the blood glucose sensor records every 5 min, the constant 288 is the number of blood glucose records in one day Finally, we use the mean values of G_{today} and G_{st} to impute the missing values with g_{i-1} and g_{i-2}.

3.4. Temporal multi-head attention model to obtain global and local spatiotemporal features

Temporal series task modeling requires the realization of a holistic perception of time series data. The temporal convolutional network can model sequential tasks, but its ability to find holes in the details of the data is weak, and it is easy to lose the useful information contained in the details. Multi-head attention [39] can selectively focus on important information. Therefore, in order to obtain global features of sequence data, and to achieve strong correlation between data. We incorporate a multi-head attention mechanism into the fully convolutional structure. However, sequence features between sequence data are very important for continuous prediction of future blood glucose values. We note that BiLSTM [40] has advantages in retaining the sequence information between sequence data. Inspired by the above methods, we propose temporal multi-head attention model to obtain global and local spatiotemporal features from sequence data, in which a temporal series features module is designed to insight the detail feature so as to keep the useful information contained in the details, and an intensity correlation module is used to obtain global features of sequence data and then to achieve strong correlation between blood glucose data, and a sequence features module is employed to learn temporal sequence features between sequence data to guarantee continuous prediction of future blood glucose values.

3.4.1. Temporal series features module

The temporal series features module uses causal convolution to memorize past information and learns more feature information through the expanded expansion convolution structure. The residual connection is also used to avoid vanishing and exploding gradients, and thus obtains deeper timing features [41] to realize the overall perception of sequence data. Fig. 2 shows a schematic of the expanded convolutional layer

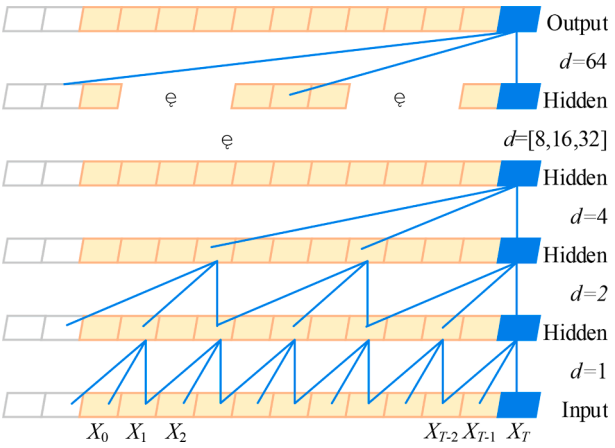


Fig. 2. Expanded convolutional layer structure.

structure. The formula for each layer of the expanded convolution $F(s)$ is as follows:

$$F(s) = (x*f)(s) = \sum_{i=0}^{k-1} f(i) \bullet x_{s-d\bullet i} \tag{1}$$

where * is the convolution operation; k is the size of the convolution kernel; d is the expansion coefficient, which has an exponential relationship with the number of network layers; and $s - d \bullet i$ is the index to the previous time–frequency information.

To improve convergence speed and reduce training time, for the learning parameter W_l^t at time t of the lth convolutional layer, we introduce the following normalized update calculation formula:

$$W_l^t = \frac{\|W_l^{t-1}\|}{\|V_{l-1}\|} V_{l-1} \tag{2}$$

where $\|W_l^{t-1}\|$ is the modulo of the learning parameters at moment $t-1$ of the l th convolutional layer, and V_{l-1} is the output vector of layer $l-1$.

The residual block and residual connection together constitute the residual unit. The structure of the residual block is shown in Fig. 3. The output feature X_n of the n th layer residual unit is as follows:

$$X_n = X_{n-1} + W_n * \delta(X_{n-1}) \tag{3}$$

where X_{n-1} denotes the output features of the residual unit at layer $n-1$, σ denotes the ReLU activation function, and W_n denotes the learnable parameters at layer n.

3.4.2. Intensity correlation module

The intensity correlation module used is stacked by h basic units of scaled dot-product attention [39]. Fig. 4 shows a schematic of the linear change calculation process of the attention mechanism module. The

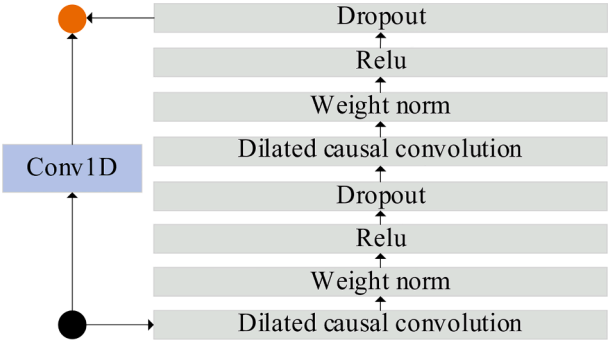


Fig. 3. Structure of residual block.

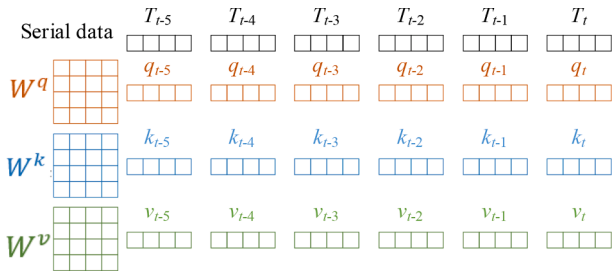


Fig. 4. Linear change calculation process of the attention mechanism module.

Attention (Q, K, V) for each layer can be calculated as

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

where Q, K, and V are the output vectors T of the fully convolutional neural network module multiplied by W^q, W^k, and W^v, respectively. d_k is the number of hidden units of the neural network.

The attention of the i th space $Head_i$ is calculated as

$$Head_i = Attention\left(W_i^Q Q, W_i^K K, W_i^V V\right) \tag{5}$$

where $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$ and $W_i^V \in \mathbb{R}^{d_{model} \times d_k}$ are the parameter matrix required for the linear mapping of the Q, K, and V matrices in the ith space, respectively.

After the attention of each space is obtained, the multi-space attention $MultiHead(Q, K, V)$ stitching calculation formula is as follows:

$$MultiHead(Q, K, V) = Concat(Head_1, Head_2, \cdots, Head_h)\, W^* \tag{6}$$

where $W^* \in \mathbb{R}^{d_{model} * d_k}$ is the weight of linear transformation.

3.4.3. Sequence features module

After the important features of the data are extracted by the intensity correlation module, the sequence features module is used to achieve bi-directional feature learning.

The intensity correlation module is based on the basic BiLSTM by splicing the two front and back LSTMs. Its basic unit is composed of the forget gate f_t, input gate i_t, and output gate O_t. The gate structure is a feedforward neural network with the sigmoid function as the activation function.

The update formula of the state of the basic unit at each time step is as follows [42]:

$$f_t = sigmoid\left(w_{xf} x_t + w_{hf} h_{t-1} + b_f\right) \tag{7}$$

$$i_t = sigmoid(w_{xi} x_t + w_{hi} h_{t-1} + b_i) \tag{8}$$

$$o_t = sigmoid(w_{xo} x_t + w_{ho} h_{t-1} + b_o) \tag{9}$$

$$c_t = f_t c_{t-1} + i_t \tanh(w_{xc} x_t + w_{xc} h_{t-1} + b_c) \tag{10}$$

$$h_t = o_t \tanh(c_t) \tag{11}$$

where x_t and h_t are the input and output at time t, sigmoid () and tanh () are activation function, and w_{xf}, w_{hf}, w_{xi}, w_{hi}, w_{xo}, and w_{ho} are the weight coefficients of the input information at the current moment and the output information at the previous moment, respectively. b_f, b_i, b_o, and b_c are the bias corresponding to the gate and cell state.

3.5. Pseudocode of PBGTAM

Algorithm 3 is the pseudocode of PBGTAM. In Step 1, the input information I_{Rate} is a combination of basal and temp_basal. To extract long-distance data features and avoid the problem of vanishing gradient, in

Step 4, we set the sliding window of the time series to w and use the padding layer to achieve causal convolution. To realize interval sampling, the expansion convolution coefficients of each layer from the first layer are set to 1, 2, 4, 8, 16, 32, and 64, respectively. The input and output dimensions of each layer are consistent through the one-dimensional CNN and then connected with the upper layer. To mine the useful information contained in the data and avoid paying too much attention to their own location, in Step 5, we set h parallel self-attention modules to build the attention layer and then the output results are respectively multiplied with the h linear transformation matrices W^q, W^k, and W^v. Compared with a single self-attention, its dimension becomes the original $1/h$.

Algorithm 3. Pseudocode of PBGTAM

Input: Insulin rate I_{Rate}, Insulin dose I_{Dose}, Carbohydrate intake C_{Input}, Continuous blood glucose level G

Output: Blood glucose prediction value G^*

1 Use Algorithm 1 to detect continuous blood glucose data and Algorithm 2 to fill in the missing values.
2 Data cleaning for I_{Rate}, I_{Dose}, and C_{Input}.
3 Preprocess the blood glucose data G and then splice these with I_{Rate}, I_{Dose}, and C_{Input} to obtain X.
4 Initialize model parameters.
5 Input X into the stacked residual module. Use Equation (1) to perform convolution operation to output the feature vector f, Equation (2) to normalize the weight, and Equation (3) to splice the residual module end to end and output the global feature vector F.
6 Initialize $Q = K = V$ and linearly change to obtain Q_i, K_i, V_i. Use Equation (4) to calculate the similarity weight matrix of the basic feature set f in a single space and multiply the weight and eigenvalue to obtain the attention vector H_i.
7 Extend Step 5 to h spaces and use Formula (6) to perform a weighted summation to obtain the global attention vector MH.
8 Input MH into the *BiLSTM* and obtain the forward *LSTM* output value $\overrightarrow{h}$ and backward *LSTM* output value $\overleftarrow{h}$ using Formulas (7) – (11). The updated hidden layer state sequence H is obtained after merging.
9 H outputs the prediction result G through the *Dense* layer and *Dropout* layer.

4. Experimental Results and Analysis

4.1. Datasets

The OhioT1DM Dataset published by Ohio University is used as the experimental data [43]. We use OhioT1DM Dataset for three main reasons. Firstly, the volume of OhioT1DM Dataset data is large enough to support our experiment, the data volume meets the requirements. Secondly, the dataset comes from 12 real patients with difference, and it also meets the requirements in terms of authenticity and generality. Finally, OhioT1DM Dataset is recognized widely in the field of blood glucose prediction. The dataset consists of glucose level, finger stick, bolus, basal, temp_basal, and meal.

OhioT1DM Database includes two types of datasets: train and test. Its data are composed of diabetics' data for eight weeks, including glucose level, finger stick, bolus, basal, temp_basal, meal, exercise, and basis heart rate. For a single patient's blood glucose dataset, the document is divided into train and test datasets. When the model is trained, the dataset in the train document is divided into the training and validation datasets according to a 9:1 ratio. When testing, we directly use the dataset of the test document as the test dataset. Table 3 shows the specific distribution plan of the dataset.

Table 3
Statistical information of the training dataset, validation dataset, and test dataset.

Patient ID	Gender	Age	Train	Validation	Test	Experiment duration/day
559	female	40–60	9716	1080	2514	40
563	male	40–60	10,912	1212	2570	40
570	male	40–60	9883	1099	2745	39
575	female	40–60	10,679	1187	2590	44
588	female	40–60	11,376	1264	2791	45
591	female	40–60	9762	1085	2760	44

4.2. Performance metrics

4.2.1. RMSE and MARD

We use root mean square error (RMSE) and mean absolute relative difference (MARD) to examine the overall prediction performance. The RMSE mainly investigates the generalizability and robustness of the model. The MARD reflects the prediction accuracy of the algorithm. The smaller the RMSE and MARD values, the better the model performance.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}(k|k-PH) - y(k)\right)^2}$$

$$MARD = \frac{1}{N}\sum_{k=1}^{N}\frac{|\hat{y}(k|k-PH) - y(k)|}{y(k)} \times 100\%$$

where N represents the number of data samples,k represents the time stamp, $y(k)$ is the blood glucose value at time k, and $\hat{y}(k|k-PH)$ is the predicted blood glucose value at the PH time in advance.

4.2.2. Forecasting accuracy for warning

We use forecasting accuracy A_f to evaluate the model's early warning ability against hyperglycemia and hypoglycemia. We then divide the blood glucose level into three levels, namely, higher than 180 mg/dl for hyperglycemia, lower than 70 mg/dl for hypoglycemia, and between 70 and 180 mg/dl for normal level. Table 4 shows the judgment matrix of the predicted grade obtained by the algorithm and the true standard grade. Therefore, the formula for calculating the forecasting accuracy A_f is as follows:

$$A_f = (n_{HH} + n_{NN} + n_{LL}) / (n_{HH} + n_{NH} + n_{LH} + n_{HN} + n_{NN} + n_{LN} + n_{HL} + n_{NL} + n_{LL})$$

where n_{HH}, n_{NN}, and n_{LL} are the statistical numbers of the prediction level of blood glucose to be the real levels of sample on different conditions, and n_{HN}, n_{HL}, n_{NH}, n_{NL}, n_{LH}, and n_{LN} are the statistical numbers of the prediction level of blood glucose to be the wrong level of sample.

4.2.3. Clark error grid analysis

The Clark error grid analysis [44] method can be used to assess the agreement between predicted and reference values of blood glucose concentrations. The main factors to be considered include: the absolute value of the predicted blood glucose value, the absolute value of the blood glucose reference value, the relative deviation of the absolute value of the predicted blood glucose value from the absolute value of the blood glucose reference value, and the clinical significance of the deviation and the resulting clinical significance.

4.3. Baseline and experiment setting

To examine the superiority of the PBGTAM algorithm, we compare it with the following six methods: (1) BiLSTM-Based [12], (2) TCN-based Method [25], (3) DRNN [35], (4) Deep-Ensemble Learning [45], (5) Modified WaveNet [46], (6) XGBoost PCA-reduced [47].

Using Keras and the TensorFlow deep learning framework, we

Table 4
Judgment matrix.

		Prediction Sample		
		High Level	Normal Level	Low Level
Actual Sample	High Level	n_{HH}	n_{HN}	n_{HL}
	Normal Level	n_{HN}	n_{NN}	n_{NL}
	Low level	n_{LH}	n_{LN}	n_{LL}

Table 5
metrics of different algorithms under 30 min and 60 min in advance for different patients.

Algorithm	Patient ID	Time = 30 min			Time = 60 min		
		RMSE	*MARD* /%	A_f /%	*RMSE*	*MARD* /%	A_f /%
TCN [25]	559	27.18	11.12	76.54	39.58	17.93	70.88
	563	20.58	8.75	79.35	37.02	15.87	70.03
	570	21.03	7.38	92.44	33.04	11.98	88.02
	575	27.93	12.16	77.11	43.04	19.27	70.28
	588	27.44	11.75	75.57	39.22	16.65	79.45
	591	24.15	10.12	75.24	41.21	15.23	74.12
	Mean	24.72	10.21	79.38	38.85	16.16	75.46
DRNN [35]	559	23.25	10.28	78.34	38.83	18.29	71.28
	563	20.55	8.90	79.85	37.84	16.24	71.90
	570	20.33	7.11	92.26	37.29	13.47	87.44
	575	27.23	11.48	79.06	45.21	20.32	68.69
	588	24.42	10.27	79.38	39.75	15.86	70.78
	591	23.12	10.24	79.41	41.28	17.34	72.18
	Mean	23.15	9.71	81.38	40.03	16.92	73.71
Modified WaveNet [46]	559	22.48	–	–	–	–	–
	563	20.35	–	–	–	–	–
	570	18.26	–	–	–	–	–
	575	25.65	–	–	–	–	–
	588	21.69	–	–	–	–	–
	591	24.59	–	–	–	–	–
	Mean	22.17	–	–	–	–	–
Deep-Ensemble Learning [45]	559	21.67	10.52	78.19	37.17	18.13	78.70
	563	20.13	8.64	79.67	35.19	15.48	72.08
	570	20.67	7.41	92.08	31.32	11.14	88.38
	575	26.22	11.81	78.99	46.06	21.29	66.75
	588	19.84	8.29	85.03	43.14	18.45	67.36
	591	22.31	9.52	81.64	42.23	18.52	72.24
	Mean	21.81	9.36	82.60	39.18	17.17	73.36
BiLSTM [12]	559	20.62	9.46	80.64	49.76	21.40	70.24
	563	20.45	8.70	80.06	38.35	17.02	69.73
	570	18.04	6.27	92.22	31.24	11.35	89.38
	575	26.19	11.08	80.50	45.76	22.36	67.97
	588	21.77	9.37	84.56	42.88	18.59	75.60
	591	22.33	10.24	82.15	43.31	23.12	71.46
	Mean	21.57	9.19	83.36	41.88	18.97	74.06
XGBoost PCA-reduced [47]	559	21.82	–	–	–	–	–
	563	19.38	–	–	–	–	–
	570	19.61	–	–	–	–	–
	575	24.24	–	–	–	–	–
	588	22.14	–	–	–	–	–
	591	23.39	–	–	–	–	–
	Mean	21.76	–	–	–	–	–
PBGTAM	559	20.84	9.44	80.11	36.08	17.80	71.43
	563	19.58	8.43	81.04	34.26	15.04	73.45
	570	18.52	6.36	92.98	31.66	11.06	89.07
	575	24.71	10.91	80.71	40.46	18.54	71.73
	588	20.20	8.17	86.87	34.64	14.08	75.79
	591	19.57	8.78	84.36	33.46	15.12	74.32
	Mean	20.57	8.68	84.35	35.09	15.27	75.97

implemented the above five methods and PBGTAM in Python language. The TCN-based Method and the temporal series features module of PBGTAM use the same setup with 64 convolutional kernels of size 4, an expansion factor of 6, and an activation function of RELU. For the BiLSTM-Based method and the intensity correlation module, the hidden layer cell is 64 and the dropout layer parameters are all 0.3. For the method with the attention mechanism module, the number of layer heads are 8 and the number of hidden layer units are 64. In addition, we use Adam optimizer and MAPE loss function, and set the number of iterations for all experiments to 450, the batch parameter to 1024, the learning rate to 0.001, and the weight decay rate to 0.0005. All programs run on the Dell 5810 server with the following main parameters: Intel Xeon E5-2630 processor, 64G of RAM, NVIDIA 1080Ti graphics card, and Ubuntu 18.04.5LTS operating system.

4.4. Results and analysis

Table 5 shows the statistics of each algorithm on different patient samples and advance times. Fig. 5 is the boxplots of the indicators. Specially, the results in Table 5 corresponding to algorithm Modified WaveNet and XGBoost PCA-reduced are directly quote from the original papers, and there is only present RMSE at 30 min prediction. Otherwise,

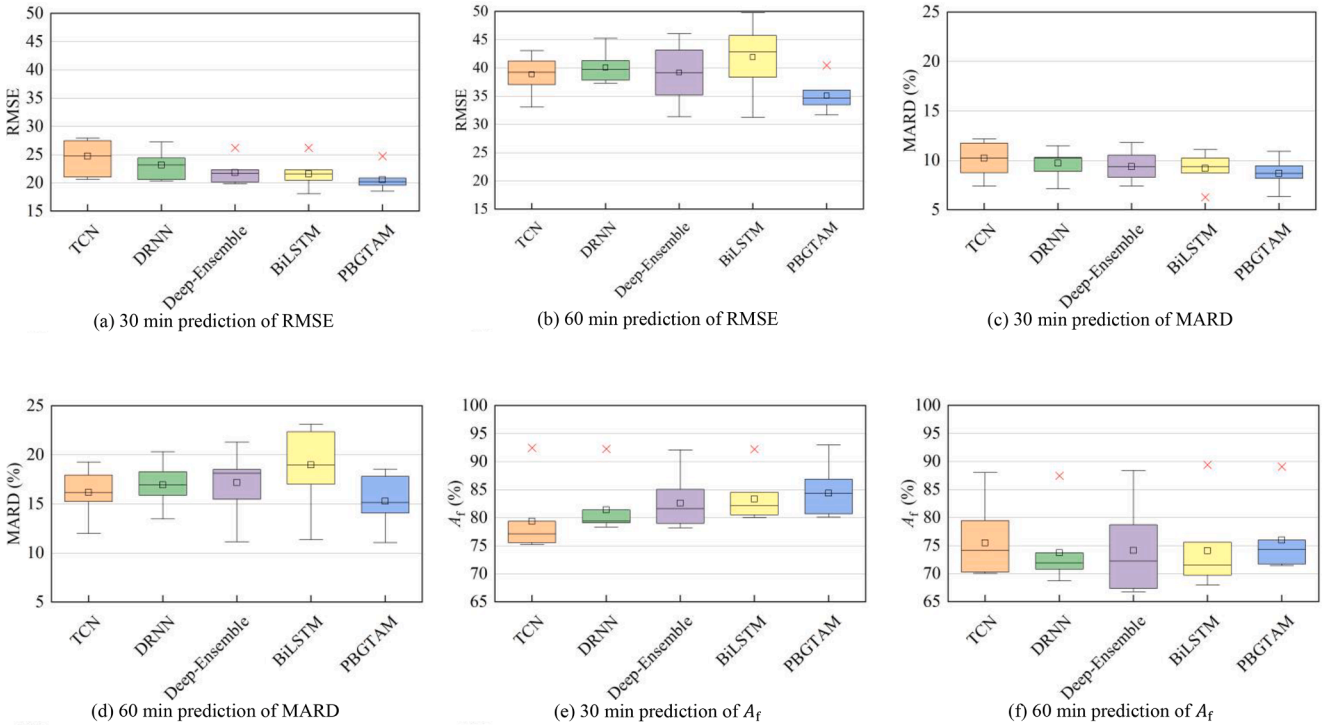


Fig. 5. Boxplots of statistic RMSE, MARD and A_f with different algorithms.

Table 6
The results of the paired sample T-test.

Algorithms	Paired Variables	Pair 1	Pair 2	Paired difference	t	P
TCN [24]	*RMSE_30*	20.57 ± 2.169	24.718 ± 3.312	−4.148 ± −1.142	−4.293	0.008***
	MARD_30	8.682 ± 1.501	10.213 ± 1.852	−1.532 ± −0.351	−3.407	0.019**
	A_f_30	84.345 ± 4.96	79.375 ± 6.564	4.97 ± −1.604	2.846	0.036**
	RMSE_60	35.093 ± 3.004	38.852 ± 3.49	−3.758 ± −0.486	−4.141	0.009***
	MARD_60	15.273 ± 2.695	16.155 ± 2.509	−0.882 ± 0.186	−2.403	0.061*
	A_f_60	75.965 ± 6.623	75.463 ± 7.11	0.502 ± −0.487	0.528	0.620
DRNN [34]	*RMSE_30*	20.57 ± 2.169	23.15 ± 2.568	−2.58 ± −0.399	−5.4	0.003***
	MARD_30	8.682 ± 1.501	9.713 ± 1.514	−1.032 ± −0.013	−4.026	0.010**
	A_f_30	84.345 ± 4.96	81.383 ± 5.352	2.962 ± −0.392	2.713	0.042**
	RMSE_60	35.093 ± 3.004	40.033 ± 2.906	−4.94 ± 0.099	−6.877	0.001***
	MARD_60	15.273 ± 2.695	16.92 ± 2.328	−1.647 ± 0.367	−5.722	0.002***
	A_f_60	75.965 ± 6.623	73.712 ± 6.838	2.253 ± −0.216	3.354	0.020**
Deep-Ensemble Learning [46]	*RMSE_30*	20.57 ± 2.169	21.807 ± 2.355	−1.237 ± −0.185	−2.687	0.043**
	MARD_30	8.682 ± 1.501	9.365 ± 1.602	−0.683 ± −0.101	−3.982	0.011**
	A_f_30	84.345 ± 4.96	82.6 ± 5.249	1.745 ± −0.289	7.035	0.001***
	RMSE_60	35.093 ± 3.004	39.185 ± 5.551	−4.092 ± −2.546	−2.471	0.056*
	MARD_60	15.273 ± 2.695	17.168 ± 3.481	−1.895 ± −0.786	−2.519	0.053*
	A_f_60	75.965 ± 6.623	74.252 ± 8.148	1.713 ± −1.525	0.799	0.460
BiLSTM [12]	*RMSE_30*	20.57 ± 2.169	21.567 ± 2.705	−0.997 ± −0.535	−2.014	0.100
	MARD_30	8.682 ± 1.501	9.187 ± 1.646	−0.505 ± −0.145	−1.886	0.118
	A_f_30	84.345 ± 4.96	83.355 ± 4.645	0.99 ± 0.315	2.177	0.081*
	RMSE_60	35.093 ± 3.004	41.883 ± 6.417	−6.79 ± −3.412	−3.389	0.019**
	MARD_60	15.273 ± 2.695	18.973 ± 4.396	−3.7 ± −1.701	−3.488	0.018**
	A_f_60	75.965 ± 6.623	74.063 ± 7.928	1.902 ± −1.305	2.603	0.048**

Note: ***, ** and * represent the significance level of 1%, 5% and 10% respectively.

the source codes of algorithms [46] and [47] are not publicly available, that's why there is no other metric corresponding to them in Table 5, and they are only the statistic results of TCN, DRNN, Deep-Ensemble Learning and BiLSTM in Tables 6-8 and Figs. 5-7.

The observations in Fig. 5(a) and 5(b) and Table 5 indicate that the RMSE of PBGTAM is less than those of TCN, DRNN, Deep-Ensemble Learning, BiLSTM. When predicting blood glucose 30 min in advance, the mean RMSE of PBGTAM is lower than those of TCN, DRNN, Deep-Ensemble Learning and BiLSTM by 4.15, 2.58, 1.24 and 1, respectively. At the same time, the average RMSE of PBGTAM decreased by 3.76, 4.94, 4.09 and 6.79 compared respectively with the above methods when blood glucose is predicted 60 min in advance.

Table 7
Clarke error grid statistics table under different methods.

Algorithms	Patient ID	30 min in advance					60 min in advance				
		A(%)	B(%)	C(%)	D(%)	E(%)	A(%)	B(%)	C(%)	D(%)	E(%)
TCN [24]	559	75.39	24.00	0.32	0.29	0.00	56.82	41.56	1.26	0.32	0.04
	563	81.96	17.78	0.22	0.04	0.00	60.06	38.54	1.30	0.10	0.00
	570	92.26	7.70	0.04	0.00	0.00	79.42	19.68	0.90	0.00	0.00
	575	76.14	22.78	0.61	0.40	0.07	59.73	36.09	2.77	1.23	0.18
	588	79.85	20.08	0.07	0.00	0.00	75.93	23.10	0.79	0.18	0.00
	591	77.51	22.35	0.14	0.00	0.00	74.78	24.87	0.36	0.00	0.00
	Mean	80.52	19.11	0.23	0.12	0.01	67.79	30.64	1.23	0.31	0.04
DRNN [34]	559	77.69	21.66	0.07	0.58	0.00	58.44	40.37	0.54	0.65	0.00
	563	81.32	18.46	0.22	0.00	0.00	58.91	39.26	1.80	0.04	0.00
	570	92.95	7.02	0.04	0.00	0.00	73.59	25.33	1.01	0.07	0.00
	575	77.26	21.16	0.54	0.83	0.22	58.01	36.96	2.45	2.16	0.43
	588	84.63	15.33	0.00	0.00	0.04	64.34	33.43	2.20	0.04	0.00
	591	90.39	9.03	90.39	0.22	0.07	65.74	29.94	65.74	1.48	0.22
	Mean	84.04	15.44	15.21	0.27	0.05	63.17	34.21	12.29	0.74	0.11
Deep-Ensemble Learning [46]	559	78.19	21.70	0.11	0.00	0.00	67.51	31.52	0.25	0.61	0.11
	563	82.66	17.09	0.22	0.04	0.00	62.18	36.88	0.94	0.00	0.00
	570	92.66	7.30	0.04	0.00	0.00	81.36	18.06	0.58	0.00	0.00
	575	77.87	21.19	0.43	0.36	0.14	53.65	42.25	2.30	1.33	0.47
	588	89.89	10.11	0.00	0.00	0.00	47.68	50.23	2.09	0.00	0.00
	591	80.68	18.86	0.25	0.22	0.00	59.77	38.00	1.87	0.36	0.00
	Mean	83.66	16.04	0.17	0.10	0.02	65.12	33.28	1.17	0.34	0.09
BiLSTM [12]	559	80.53	19.04	0.07	0.36	0.00	51.95	45.23	2.20	0.58	0.04
	563	82.01	17.70	0.22	0.07	0.00	58.26	40.48	1.26	0.00	0.00
	570	93.66	6.30	0.04	0.00	0.00	80.24	18.78	0.94	0.04	0.00
	575	78.63	20.15	0.54	0.50	0.18	52.88	42.61	2.20	2.09	0.22
	588	87.33	12.63	0.00	0.04	0.00	71.35	27.17	1.48	0.00	0.00
	591	85.17	14.75	0.00	0.07	0.00	66.71	29.58	2.09	1.51	0.11
	Mean	84.56	15.10	0.14	0.17	0.03	63.57	33.98	1.69	0.70	0.06
PBGTAM	559	80.10	19.83	0.07	0.00	0.00	60.63	38.18	0.29	0.90	0.00
	563	83.34	16.48	0.18	0.00	0.00	63.58	35.48	0.94	0.00	0.00
	570	93.59	6.37	0.04	0.00	0.00	81.32	17.70	0.94	0.04	0.00
	575	78.81	20.01	0.47	0.65	0.07	58.44	38.14	1.80	1.40	0.22
	588	89.06	10.90	0.04	0.00	0.00	71.07	27.89	0.86	0.18	0.00
	591	86.18	12.56	0.43	0.83	0.00	72.50	25.95	1.12	0.43	0.00
	Mean	85.18	14.36	0.20	0.25	0.01	67.92	30.56	0.99	0.49	0.04

Table 8
The computational cost of different Algorithms.

Algorithms	FLOPs/ G
TCN [25]	3.8e−03
DRNN [35]	7.87e−06
Deep-Ensemble Learning [45]	3.82e−03
BiLSTM [12]	4.95e−05
PBGTAM	5.09e−03

From a microscopic point of view, the RMSE of the PBGTAM algorithm in test 559, 563, 570, 575, 588 and 591 are 20.84, 19.58, 18.52, 24.71, 20.20 and 19.57, respectively. The maximum deviation of the RMSE value between different patients is only 6.19, which is relative to 7.35 of TCN, 6.9 of DRNN, 7.39 of Modified WaveNet, 6.38 of Deep-Ensemble Learning,8.15 of BiLSTM and 4.86 of XGBoost PCA-reduced. Except XGBoost PCA-reduced, the RMSE deviation value of PBGTAM decreased by 1.16, 0.71, 1.2, 0.19 and 1.96, respectively. Similarly, except DRNN, Modified WaveNet and XGBoost PCA-reduced, in 60 min blood glucose prediction, the RMSE deviation value of PBGTAM is also lower than the other three comparison algorithms by 1.2, 5.94, and 9.72, respectively. On the whole, which reveals that the PBGTAM algorithm has a relative robustness.

As observed in Fig. 5(c) and 5(d), the overall error between the predicted and true blood glucose values of PBGTAM is smaller compared to those of other algorithms. According to Table 5, we can see that the average error of PBGTAM is lower than those of TCN, DRNN, Deep-Ensemble Learning and BiLSTM by 1.53 %, 1.03 %, 0.68 % and 0.51 %, respectively, at the 30 min prediction. In the 60 min prediction, the average error of PBGTAM decreased by 0.89 %, 1.65 %, 1.9 % and 3.7 %, respectively, compared with the above four methods. Meanwhile, which illustrates that the overall trend of the predicted values of PBGTAM is closer to the real values and shows that PBGTAM has stronger prediction accuracy than the other methods.

As can be seen from Fig. 5(e) and 5(f), the PBGTAM algorithm has high accuracy in blood glucose early warning. Moreover, Table 5 shows that in 30 min blood glucose prediction, the average forecasting accuracy for warning (A_f) of the PBGTAM algorithm is 0.6 % higher than that of the other best performing BiLSTM, and the average A_f is 4.97 % higher than that of the worst performing TCN. And at the 60 min glucose prediction, the PBGTAM algorithm improved A_f by 1.91 % over BiLSTM and 0.51 % over TCN.

To further clarify whether our proposed PBGTAM method has a significant performance improvement over compared methods, we used a paired sample T-test to compare PBGTAM with other methods in pairs. The null hypothesis of paired sample T-test is that the data obtained by PAGTAM method is not significantly different from that obtained by other methods. Table 6 shows the results are significant (20/4), which means that the PAGTAM method is significantly superior to other methods.

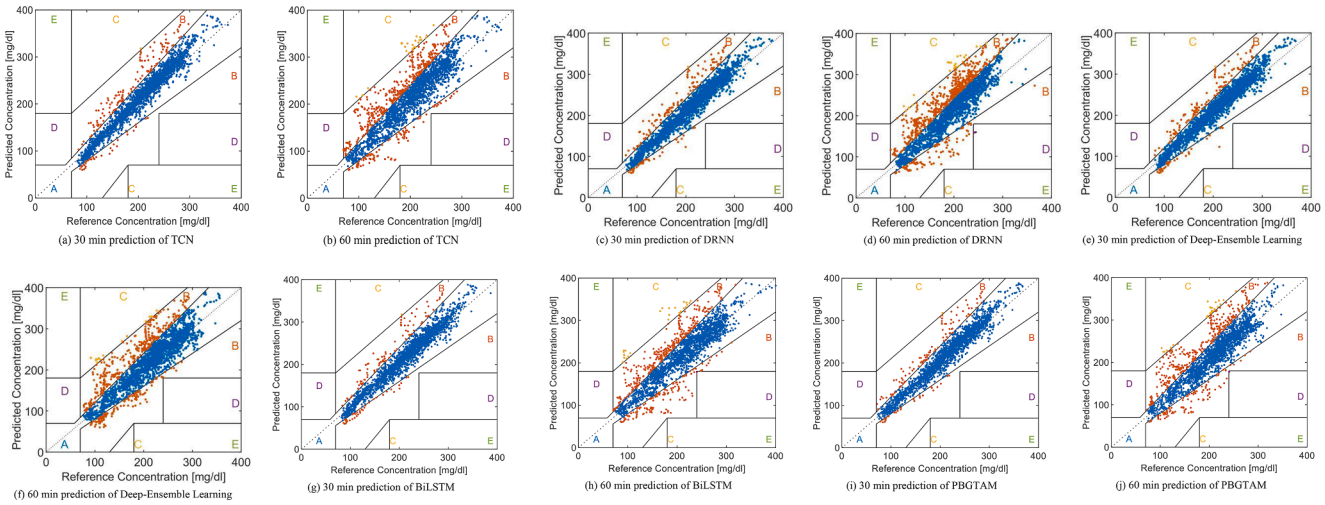


Fig. 6. Clarke error grid of different algorithms with a 30 min and 60 min prediction on patient 570.

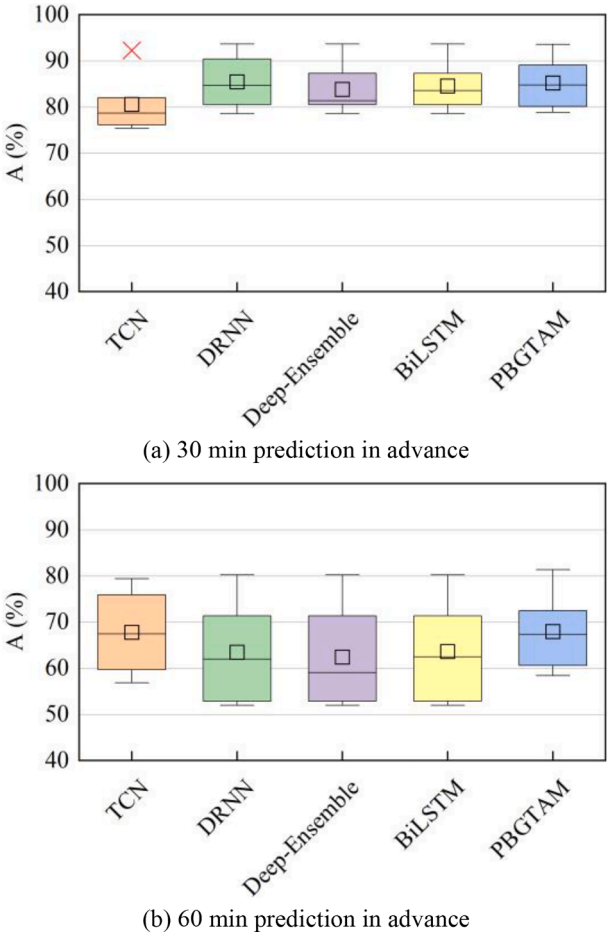


Fig. 7. Boxplots of the proportion of scattered points falling in area A with different algorithms.

Clarke grid error analysis (Fig. 6) is a scatter diagram of the correlation between the predicted value and the real value of patient 570, in which the horizontal axis reading corresponding to each point in the figure is the target blood glucose values, and the vertical axis is the blood glucose predicted values of the algorithm in this paper. When the point is closer to the diagonal of the graph, it indicates that the predicted value is closer to the true value, that is, the higher accuracy.

Table 7 is the Clarke error grid statistics table on different algorithms, and Fig. 7 is the boxplot of the proportion of scattered points of each algorithm in the A area.

Fig. 7 reveals that the red line of the PBGTAM algorithm is higher than that of other algorithms, which indicates that the proportion of scattered points falling in area A is at a relatively high level, that is, the predicted values of PBGTAM and the true blood glucose values are clinically consistent.

At the same time, the result can be found in Table 7 as follows.

- When predicting blood glucose 30 min in advance, with the exception of 575, the prediction data of the remaining patients fall in the acceptable A area, which accounts for more than 80 % of the whole. According to the observation in Fig. 6(i), the data of patient 570 fall in the A area, accounts for 93.59 %. Occasionally, a pair of points fall in the B area, which is a slight overestimation or underestimation area, and almost no points fall in areas C, D, and E, which have obvious errors. On the contrary, the scatters of TCN, DRNN, Deep-Ensemble Learning and BiLSTM algorithms fall in a larger proportion in region B and less in region A. The average number of scatters falling in region A of the PBGTAM algorithm has a higher percentage of scatters than that of TCN, DRNN, Deep-Ensemble Learning and BiLSTM, which improved by 4.66 %, 1.14 %, 1.52 % and 0.62 %, respectively.
- When predicting blood glucose 60 min in advance, the PBGTAM algorithm falls in both the A and B areas combined with a percentage above 95 % for all patients, which is an acceptable accuracy for clinical guidance [48], and except patient 575, the number falling in area A accounts for more than 60 %. According to Fig. 6(j), the number of scattered points in data 570 that fall in the A area accounts for more than 80 %, and very few scattered points fall in the obviously wrong areas C, D, and E. Compared with other algorithms, the proportion of the average number of scattered points in the A area of the PBGTAM algorithm increased by 0.13 %, 4.75 %, 2.8 % and 4.35 %, respectively.

Otherwise, focus on the computational efficiency, we counted the FLOPs of each model. The statistical results are shown in Table 8. According to Table 8, it can be found that PBGTAM has the highest FLPOs, which means that the performance improvement also along with the increase of computation when using the multi-head structure.

In summary, PBGTAM has better blood glucose prediction ability in line with clinical manifestations than the six compared methods, and is superior to other algorithms in generalizability, robustness, prediction accuracy, and forecasting accuracy for warning.

5. Conclusions and discussion

Accurate blood glucose prediction plays an important role in the future blood glucose management of patients, and the quality of blood glucose data and prediction models often affect the accuracy of the prediction results. In field of diabetes, patients need regular and frequent needle and blood sampling in blood glucose measurement. Accurate blood glucose prediction can reduce the frequency of blood sampling, thereby reducing the pain caused by blood glucose measurement to patients. However, the data obtained by continuous blood glucose data acquisition equipment are often abnormal or even lost, yet existing methods based on TCN or LSTM cannot obtain the global and temporal features of sequence data at the same time, which brings challenges to blood glucose prediction.

To meet this challenge, this study proposes the detection algorithm of abnormal blood glucose based on ADBSCAN, designs a blood glucose missing values imputation algorithm based on feature engineering, and constructs a blood glucose prediction model based on temporal multi-head attention mechanism. The ADBSCAN algorithm without prior knowledge can adaptively generate different neighborhood parameters according to the different constitutions of each patient. In this way, outliers in blood glucose data can be effectively detected to ensure the quality of prediction. The proposed PBGTAM algorithm can extract time series features, establish the intensity correlation between blood glucose data, and obtain time series information of blood glucose data, which is superior to the six compared algorithms in generalizability, robustness, prediction accuracy, and forecasting accuracy for warning. It also has better blood glucose prediction ability in line with clinical manifestations than the other six methods. Especially when predicting blood glucose 60 min in advance, PBGTAM has more significant advantages in each index.

However, PBGTAM still has some limitations.

- It has no advantage in computational complexity compared to the model being compared. Thus, we will refine PBGTAM in the future.
- For the loss of the data itself and the missing values phenomenon caused by the later discarding of outliers, the imputation algorithm based on feature engineering for missing values are used to fill the data. Although the filling effect is good, owing to the small amount of blood glucose data collected from the fingertips, the limited reference data cannot fully measure the performance of the algorithm. Therefore, it is meaningful work to construct a real dataset for the United Hospital in later research.
- Even though our data come from real patients, since the age of the selected test data is concentrated between 40 and 60 years and the data diversity are not strong, how to build a diverse dataset to improve the long-term prediction ability and robustness of PBGTAM will be the focus in future.
- The proposed method has not been applied in clinical practice, and the situation of actual patients is much more complex than that contained in the test dataset. How to improve the robustness of the model to adapt to real patients is an important work. Of course, using the proposed method to develop a clinical application system requires more in-depth research.

CRediT authorship contribution statement

Guanci Yang: Investigation, Formal analysis, Funding acquisition, Supervision, Resources, Writing – review & editing. **Saisai Liu:** Investigation, Formal analysis, Software, Writing – original draft. **Yang Li:** Formal analysis, Funding acquisition, Writing – review & editing. **Ling He:** Validation, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] H. Sun, P. Saeedi, S. Karuranga, M. Pinkepank, K. Ogurtsova, B.B. Duncan, C. Stein, A. Basit, J.C. Chan, J.C. Mbanya, IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045, Diabetes Res. Clin. PR 183 (2022), 109119.

[2] J.J. Chamberlain, A.S. Rhinehart, C.F. Shaefer Jr, A. Neuman, Diagnosis and management of diabetes: synopsis of the 2016 American Diabetes Association Standards of Medical Care in Diabetes, Ann. Intern. Med. 164 (2016) 542-552.

[3] Y. Zhang, J. Sun, L. Liu, H. Qiao, A review of biosensor technology and algorithms for glucose monitoring, J. Diabetes Complicat. 35 (2021), 107929.

[4] Z. Nie, M. Rong, K. Li, Blood glucose prediction based on imagingphotoplethysmography in combination with Machine learning, Biomed. Signal Process. 79 (2023), 104179.

[5] S. Oviedo, J. Vehí, R. Calm, J. Armengol, A review of personalized blood glucose prediction strategies for T1DM patients, Int. J. Numer. Meth. Bio. 33 (2017) e2833.

[6] M. Munoz-Organero, Deep physiological model for blood glucose prediction in T1DM patients, Sens.-Basel 20 (2020) 3896.

[7] A. Aliberti, A. Bagatin, A. Acquaviva, E. Macii, E. Patti, Data driven patient-specialized neural networks for blood glucose prediction, 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE, 2020, pp. 1-6.

[8] S. Mirshekarian, R. Bunescu, C. Marling, F. Schwartz, Using LSTMs to learn physiological models of blood glucose behavior, 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017, pp. 2887-2891.

[9] H.N. Mhaskar, S.V. Pereverzyev, M.D. Van der Walt, A deep learning approach to diabetic blood glucose prediction, Front. Appl. Math. Stat. 3 (2017) 14.

[10] S. Naveena, A. Bharathi, A new design of diabetes detection and glucose level prediction using moth flame-based crow search deep learning, Biomed. Signal Process. 77 (2022), 103748.

[11] J. Bi, X. Zhang, H. Yuan, J. Zhang, M. Zhou, A hybrid prediction method for realistic network traffic with temporal convolutional network and lstm, IEEE Trans. Autom. Sci. Eng., (2021).

[12] Q. Sun, M.V. Jankovic, L. Bally, S.G. Mougiakakou, Predicting blood glucose with an lstm and bi-lstm based deep neural network, 2018 14th symposium on neural networks and applications (NEUREL), IEEE, 2018, pp. 1-5.

[13] Y. Wang, G. Yang, S. Li, Y. Li, L. He, D. Liu, Arrhythmia classification algorithm based on multi-head self-attention mechanism, Biomed. Signal Process. 79 (2023), 104206.

[14] Z. Jia, L. Huang, H. Liu, Y. Huang, W. Li, X. Pi, X. Zheng, Design of a real-time self-adjusting calibration algorithm to improve the accuracy of continuous blood glucose monitoring, Appl. Biochem. Biotech. 190 (2020) 1163–1176.

[15] L. Meneghetti, A. Facchinetti, S. Del Favero, Model-based detection and classification of insulin pump faults and missed meal announcements in artificial pancreas systems for type 1 diabetes therapy, IEEE Trans. Bio-Med. Eng. 68 (2020) 170–180.

[16] X. Yu, X. Sun, Y. Zhao, J. Liu, H. Li, Fault detection of continuous glucose measurements based on modified k-medoids clustering algorithm, Neural Comput. & Appl. (2020) 1–10.

[17] Y. Zhou, W. Zhu, Q. Fang, L. Bai, Survey of Outlier Detection Methods Based on Clustering, Comput. Eng. Appl. 57 (2021) 37–45.

[18] Y. Li, G. Yang, Z. Su, S. Li, Y. Wang, Human activity recognition based on multienvironment sensor data, Inform. Fusion 91 (2023) 47–63.

[19] M. Breton, B. Kovatchev, Analysis, modeling, and simulation of the accuracy of continuous glucose sensors, J. Diabetes Sci. Technol. 2 (2008) 853–862.

[20] D.J. Lunn, C. Wei, R. Hovorka, Fitting dynamic models with forcing functions: application to continuous glucose monitoring in insulin therapy, Stat. Med. 30 (2011) 2234–2250.

[21] A. Facchinetti, S. Del Favero, G. Sparacino, C. Cobelli, Model of glucose sensor error components: identification and assessment for new Dexcom G4 generation devices, Med. Biol. Eng. Compu. 53 (2015) 1259–1269.

[22] M. Vettoretti, C. Battocchio, G. Sparacino, A. Facchinetti, Development of an error model for a factory-calibrated continuous glucose monitoring sensor with 10-day lifetime, Sens.-Basel 19 (2019) 5320.

[23] Q. Shen, S.J. Qin, K.J. Doniger, Online dropout detection in subcutaneously implanted continuous glucose monitoring, Proceedings of the 2010 American Control Conference, IEEE, 2010, pp. 4373-4378.

[24] M.F. Rabby, Y. Tu, M.I. Hossen, I. Lee, A.S. Maida, X. Hei, Stacked LSTM based deep recurrent neural network with kalman smoothing for blood glucose prediction, BMC Med. Inform. Decis. 21 (2021) 1–15.

[25] J. Xie, Q. Wang, Benchmarking machine learning algorithms on blood glucose prediction for type I diabetes in comparison with classical time-series models, IEEE Trans. Bio-Med. Eng. 67 (2020) 3101–3124.

[26] K. Turksoy, L. Quinn, E. Littlejohn, A. Cinar, Monitoring and fault detection of continuous glucose sensor measurements, 2015 American Control Conference (ACC), IEEE, 2015, pp. 5091-5096.

[27] S. Lim, S. Lee, S. Kim, Clustering of detected targets using DBSCAN in automotive radar systems, 2018 19th international radar symposium (IRS), IEEE, 2018, pp. 1-7.
[28] M. Eren-Oruklu, A. Cinar, L. Quinn, D. Smith, Adaptive control strategy for regulation of blood glucose levels in patients with type 1 diabetes, J. Process. Contr. 19 (2009) 1333–1346.
[29] A.Z. Woldaregay, E. Årsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, G. Hartvigsen, Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes, Artif. Intell. Med. 98 (2019) 109–134.
[30] E.I. Georga, J.C. Príncipe, D.I. Fotiadis, Short-term prediction of glucose in type 1 diabetes using kernel adaptive filters, Med. Biol. Eng. Compu. 57 (2019) 27–46.
[31] Q. Wang, P. Molenaar, S. Harsh, K. Freeman, J. Xie, C. Gold, M. Rovine, J. Ulbrecht, Personalized state-space modeling of glucose dynamics for type 1 diabetes using continuously monitored glucose, insulin dose, and meal intake: an extended Kalman filter approach, J. Diabetes Sci. Technol. 8 (2014) 331–345.
[32] W. Wang, S. Wang, X. Wang, D. Liu, Y. Geng, T. Wu, A glucose-insulin mixture model and application to short-term hypoglycemia prediction in the night time, IEEE Trans. Bio-Med. Eng. 68 (2020) 834–845.
[33] R. Bunescu, N. Struble, C. Marling, J. Shubrook, F. Schwartz, Blood glucose level prediction using physiological models and support vector regression, 2013 12th International Conference on Machine Learning and Applications, IEEE, 2013, pp. 135-140.
[34] K. Saiti, M. Macaš, L. Lhotská, K. Štechová, P. Pithová, Ensemble methods in combination with compartment models for blood glucose level prediction in type 1 diabetes mellitus, Comput. Meth. Prog. Bio 196 (2020), 105628.
[35] T. Zhu, K. Li, J. Chen, P. Herrero, P. Georgiou, Dilated recurrent neural networks for glucose forecasting in type 1 diabetes, J. Healthcare Inform. Res. 4 (2020) 308–324.
[36] K. Li, C. Liu, T. Zhu, P. Herrero, P. Georgiou, GluNet: A deep learning framework for accurate glucose forecasting, IEEE J. Biomed. Health 24 (2019) 414–423.
[37] K. Li, J. Daniels, C. Liu, P. Herrero, P. Georgiou, Convolutional recurrent neural networks for glucose prediction, IEEE J. Biomed. Health 24 (2019) 603–613.
[38] G. Dong, Y. Jin, S. Wang, W. Li, Z. Tao, S. Guo, DB-Kmeans: An Intrusion Detection Algorithm Based on DBSCAN and K-means, IEICE Proc. Series 56 (2019).
[39] N.S.N.P. Ashish Vaswani, Attention Is All You Need, Proceedings of 31st Cofference on Neural Information Processing Systems, Long Beach.USA, 2017, pp. 6000–6010.
[40] T. Chen, R. Xu, Y. He, X. Wang, Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN, Expert Syst. Appl. 72 (2017) 221–230.
[41] K.A.Z.X. He, Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778.
[42] X.S.C.H. Yong Yu, A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures, Neural Comput. 7 (2019) 1235–1270.
[43] C. Marling, R. Bunescu, The OhioT1DM Dataset for Blood Glucose Level Prediction: Update 2020, CEUR Workshop Proc 2675 (2020) 71–74.
[44] W.L. Clarke, D. Cox, L.A. Gonder-Frederick, W. Carter, S.L. Pohl, Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose, Diabetes Care 10 (1987) 622–628.
[45] N. Hoda, K. Heydar, E.M. R, E. Jackie, B. Mohammed, Blood Glucose Level Prediction: Advanced Deep-Ensemble Learning Approach., IEEE J. Biomed. Health PP (2022).
[46] T. Zhu, K. Li, P. Herrero, J. Chen, P. Georgiou, A Deep Learning Algorithm for Personalized Blood Glucose Prediction., KHD@ IJCAI, 2018, pp. 64-78.
[47] C. Midroni, P. Leimbigler, G. Baruah, M. Kolla, A. Whitehead, Y. Fossat, Predicting glycemia in type 1 diabetes patients: experiments with xg-boost, KHD@ IJCAI, 2018.
[48] S. Garg, H. Zisser, S. Schwartz, T. Bailey, R. Kaplan, S. Ellis, L. Jovanovic, Improvement in glycemic excursions with a transcutaneous, real-time continuous glucose sensor: a randomized controlled trial, Diabetes Care 1 (2007) 44–50.