Retrospective Study

High-precision blood glucose prediction and hypoglycemia warning based on the LSTM-GRU model

Xiuli Peng1, Quanzhong Li2*, Yannian Wang3 and Dengfeng Yan1

1Emergency Department, Zhoukou Central Hospital, Zhoukou, China
2Department of Endocrinology, Henan Provincial People’s Hospital, Zhengzhou, China
3School of Computer and Artificial Intelligence, Zhengzhou University, China

Abstract

Objective: The performance of blood glucose prediction and hypoglycemia warning based on the LSTM-GRU (Long Short Term Memory - Gated Recurrent Unit) model was evaluated.

Methods: The research objects were 100 patients with Diabetes Mellitus (DM) who were chosen from Henan Provincial People’s Hospital. Their continuous blood glucose curves of 72 hours were acquired by a Continuous Glucose Monitoring System (CGMS). The blood glucose levels were predicted based on the LSTM, GRU and LSTM-GRU models, respectively. Analyses of the best predictive model were performed using Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and correlation analysis between the prediction blood glucose level and the original blood glucose level acquired by CGMS and Clark Error Grid Analysis (EGA). Repeated-measures analysis of variance (ANOVA) was used to analyze whether the RMSE values of the three models were statistically significant. 60 patients who had experienced hypoglycemia among 100 cases were selected for hypoglycemia warning. The sensitivity, false-positive rate and false-negative rate were used to evaluate the hypoglycemia warning performance of the LSTM-GRU model. This paper explored the changing relationship of the hypoglycemia warning performance of the model over time.

Results: The predicted blood glucose levels of the three models were strongly correlated with the blood glucose levels acquired by CGMS (p < 0.001). The correlation coefficient (R-value) of the LSTM-GRU model remained stable over time (R = 0.995), nevertheless, a reduction in the R-value of the LSTM and GRU models when the Prediction Horizon (PH) was 30 min or longer. When PH was 15 min, 30 min, 45 min and 60 min, the mean RMSE values of the LSTM-GRU model were 0.259, 0.272, 0.275 and 0.278 (mmol/l), respectively, which were lower than the LSTM and GRU models and the RMSE values were statistically significant (p < 0.001). The EGA results showed the LSTM-GRU model had the highest proportion in zones A and B, as the PH extended. When PH was 30 min or longer, the sensitivity and false-negative rate of the hypoglycemia warning of the LSTM-GRU model had subtle changes and the false-positive rate remained stable over time.

Conclusions: The LSTM-GRU model demonstrated good performance in blood glucose prediction and hypoglycemia warning.

Introduction

Diabetes mellitus (DM) is a group of chronic metabolic conditions, all of which are characterized by elevated blood glucose levels resulting from the body’s inability to produce insulin or resistance to insulin action, or both [1]. Blood glucose levels that are too high or too low can cause a series of diabetes-related complications [2-4].
obtain complete blood glucose levels fluctuation curve by converting glucose concentration in ISF to blood glucose levels. Therefore, CGMS can be used to assess blood glucose levels comprehensively and document more accurately the actual incidence of hypoglycemia and make it possible to develop blood glucose prediction models [7-8].

Blood glucose prediction models [9] mainly include data-driven blood glucose prediction models, physiological blood glucose prediction models and mixed-blood glucose prediction models. Physiological models [10] are usually built based on extensive knowledge and understanding of insulin, glucose metabolism and other parameters. Data-driven models [11] mainly rely on blood glucose measurements. Hybrid blood glucose prediction models [12] combine the two previous approaches together. Since physiological models are somewhat time-consuming and require prior knowledge to set physiological constants, data-driven blood glucose prediction models have gained popularity in recent years. Yang, et al. [13], proposed an autoregressive integral moving average (ARIMA) model with an adaptive recognition algorithm of model for blood glucose prediction and hypoglycemia warning. Sparacino, et al. [14], demonstrated a first-order autoregressive (AR) model to predict blood glucose levels, with a prediction horizon (PH) of 45min. Wang, et al. [15], proposed a new adaptive weighted average framework for blood glucose prediction algorithms, of which the main idea was to give each algorithm an adaptive weight, where the weight of each algorithm was inversely proportional to the sum of squared prediction errors. The method achieved satisfactory results for blood glucose prediction and it had very strong robustness to changes in patients and PHs. Pérez-Gandía, et al. [16], used an Artificial Neural Network (ANN) model based on CGMS data. Fernandez, et al. [17], used an artificial neural network to predict blood glucose levels based on patient dynamics, CGMS measurements, and insulin doses. Wang, et al. [18], proposed a short-term blood glucose prediction model combining variational mode decomposition (VMD) and an improved Particle swarm optimization optimizing long short-term memory network (IPSO-LSTM) and the model had high prediction accuracy even when PH was extended to 60min. In general, the more accuracy of blood glucose prediction and the longer PH can provide clinicians and patients with sufficient time to prevent hypoglycemia events.

With the development of information technology, more and more machine learning algorithms are introduced in the field of blood glucose prediction, such as Long Short-Term Memory (LSTM) [19,20] and GRU [21]. Due to the working principles of the LSTM and GRU being similar and these two models being deficient in each other, this paper proposed a composite model for blood glucose levels prediction and hypoglycemia warning.

The structure is set as follows: In the first part, an introduction to the basic principle of the three models is carried out. The second part introduces the experimental dataset and the process followed to develop a set of indicators that are evaluated in this paper. The third part presents the experimental results. The fourth part gives the discussion. And the last part provides conclusions of the paper and outlines directions for future research.

**Modeling principle**

To predict blood glucose levels, a combination model was proposed, illustrated in Figure 1. It contains an LSTM layer, a GRU layer and a fully connected layer. Take several blood glucose values as input and train through LSTM and GRU. A fully connected layer to output the predicted level.

The first layer is the LSTM layer, illustrated in Figure 2, consisting of a forget gate $f_t$, an input gate $i_t$ and an output gate $o_t$. The forget gate decides what information to discard. The input gate determines what information is inputted to the cell state. The calculating process in LSTM cells is as follows.

- $f_t = \sigma(W_f x + h_{t-1} x_f + b_f)$
- $i_t = \sigma(W_i x + h_{t-1} x_i + b_i)$
- $\tilde{C}_t = \tanh(W_c x + h_{t-1} x_c + b_c)$
- $C_t = f_t C_{t-1} + i_t \tilde{C}_t$
- $h_t = o_t \tanh(C_t)$

**Figure 1:** Principle of the combinational model.

**Figure 2:** Principle of the LSTM.
Where the \(x_t\) is input and \(h_{t-1}\) is the hidden state of the previous moment. \(W, b\) represent the weights matrix and biases vector, respectively and \(C_{t-1}\) is the cell state of the previous moment. In addition, \(h_{t-1}\) is hidden state of the previous moment and \(C_t\) is candidate cell state.

The \(h_{t-1}\) and \(x_t\) are activated by the sigmoid function of the output gate to obtain \(o_t\) and the new cell state is activated by the tanh function to update the hidden state \(h_t\).

\[
 o_t = \sigma(W_o \cdot [h_{t-1},x_t] + b_o) \\
 h_t = o_t \cdot \tanh(C_t) \\
 W_o, W, W_{ht}, b_o, b, b_o, b_t\) are parameters of the model and can be learned.

The above content is the calculation process of the LSTM and its entire network follows the rules of backpropagation and gradient descent to update parameters. This structure can effectively screen the effective features of long-term data and solve the long-dependency problem of the RNN (recurrent neural network).

The second layer is the GRU unit, illustrated in Figure 3. The two gates of the GRU are called the reset gate \(r_t\) and the update gate \(z_t\). \(r_t\) determines how much of the secret state at the last moment is retained and how much is reset. \(z_t\) is used to control the degree to which the state information of the previous moment is brought into the current state. The updated formula is as follows.

\[
 r_t = \sigma(W_r \cdot [h_{t-1},x_t]) \\
 z_t = \sigma(W_z \cdot [h_{t-1},x_t]) \\
 \tilde{h}_t = \tanh(W_\tilde{h} \cdot [r_t \cdot h_{t-1},x_t]) \\
 h_t = (1-z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \\
 \text{Where } \tilde{h}_t \text{ is the candidate’s hidden state. } W_r, W_z, W_\tilde{h} \text{ are parameters of the model and can be learned.}

Finally, extract all hidden states, make the final output hidden state \(h_f = (h_1,h_2,...,h_t)\) and then calculate with the output weight \(W_o\) and the output bias vector \(b_o\) to obtain the predicted value \(\hat{y}\).

\[
 \hat{y} = W_o h_f + b_o 
\]

**Experiments**

**A. Dataset:** The blood glucose curves of 100 patients with DM who received subcutaneous insulin infusion therapy during hospitalization in the Endocrinology Department of Henan Provincial People’s Hospital were retrospectively analyzed from March 2017 to December 2017. All patients met the World Health Organization (WHO) DM diagnostic criteria [22]. The following DM patients were excluded: DM patients with critically ill and unstable patients, gestational diabetes, allergies, or a history of tape allergy, the wearing time is less than 72 hours and the original blood glucose levels sequence has a breakpoint.

**B. Data preprocessing:** The original CGMS data was nonlinear and non-stationary. Firstly, we decomposed the original CGMS data using the sym5 wavelet transform [23], removed high-frequency signals, kept low-frequency signals and used a one-dimensional reconstruction function to obtain denoised blood glucose curves, illustrated in Figure 4. This method improved the validity of the original CGMS data and the accuracy of the model prediction, to some extent. Secondly, To get a better prediction effect, a min-max normalization was used to transform blood glucose levels to the range \((0,1)\).

**C. Performance evaluation:** Correlation analysis was used to evaluate the degree of correlation between the predicted blood glucose level and the CGMS data. The greater the correlation, the better the prediction is. RMSE (Root Mean Square Error) was used to evaluate the accuracy of the prediction. The lower the RMSE, the better the prediction is. The RMSE was calculated as follows:

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

**Figure 3:** Principle of the GRU

**Figure 4:** Wavelet denoising.
Mean Square Error), MAPE (Mean Absolute Percentage Error) and MAE (Mean Absolute Error) [24], were used in this paper to evaluate the prediction performance of the LSTM, GRU and LSTM–GRU models. This paper took RMSE as a statistical indicator and repeated-measures analysis of variance (ANOVA) was used to compare the prediction performance between the three models and obtained the best predictive model.

The above statistical methods were all operated by the SPSS 26.0 and \( p < 0.05 \) was considered statistically significant.

**Clark Error Grid Analysis (EGA) [26]:** Using the Beckman analyzer as the reference, the grid is subdivided into five zones: A, B, C, D and E. Values in zones A and B represent accurate or acceptable and values in Zone C, D and E represent error. A PH of 15min is the most accurate, but clinicians and patients have time to adjust their treatment. This paper compared three models with different PHs: 15min, 30min, 45min and 60min.

The EGA results of the three models are shown in Table 4 and Figure 5. Due to zones C, D and E being of little significance for clinical reference, it was not listed in Table 4.

**Results**

**A. Pearson correlation analysis**

The results of Pearson correlation analysis between the predicted blood glucose level of the three models and the original blood glucose level acquired by CGMS are shown in Table 1. For different PHs, the predicted blood glucose level of each model was positively correlated with the original blood glucose level acquired by CGMS (\( p < 0.001 \)). The R – values (\( R = 0.995 \)) of the LSTM–GRU model were identical when different PHs, nevertheless, a reduction in the LSTM and GRU model's performance start to appear when PH was 30 min or longer.

**Table 1:** Correlation coefficients of the three models with different PHs.

<table>
<thead>
<tr>
<th>model</th>
<th>15min</th>
<th>30min</th>
<th>45min</th>
<th>60min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>p</td>
<td>R</td>
<td>p</td>
</tr>
<tr>
<td>LSTM–GRU</td>
<td>0.995</td>
<td>&lt;0.001</td>
<td>0.995</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.995</td>
<td>&lt;0.001</td>
<td>0.995</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>GRU</td>
<td>0.995</td>
<td>&lt;0.001</td>
<td>0.995</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

**Note:** Pearson correlation coefficient (R): 0.8–1.0 very strong correlation.

**B. Repeated-measures ANOVA**

The prediction performance of 100 DM patients is shown in Table 2 and the results of the RMSE value variance analysis are provided in Table 3. When PH was 15min, 30min, 45min and 60min, the mean RMSE values of the LSTM–GRU model were 0.259, 0.272, 0.275 and 0.278 (mmol/l), respectively. The mean RMSE values of the LSTM–GRU model were lower than those of the LSTM and GRU models with identical PHs and had statistically different (\( p < 0.001 \)), while the mean RMSE values of the LSTM and GRU models had no statistical significance (\( p > 0.05 \)).

**C. Clark error grid analysis**

The EGA results of the three models are shown in Table 4 and Figure 5. Due to zones C, D and E being of little significance for clinical reference, it was not listed in Table 4.

**D. Hypoglycemia warning**

The hypoglycemia warning performance of 60 DM patients with hypoglycemia is shown in Table 5. When PH was 15min, 30min, 45min and 60min, the mean sensitivity were 91.21%, 89.71%, 89.21% and 88.73%, the mean false-negative rate were 8.79%, 10.29%, 10.79% and 11.27%, the mean false-positive rate were 0.88%, 0.90%, 0.87% and 0.87%. As the PH was extended, the results of the false-positive rate were almost identical, nevertheless, the sensitivity of the model gradually decreased, but the false-negative rate gradually increased and some differences in the model performance start to appear when the PH was 30min. The results of the statistical analysis of sensitivity are shown in Table 6. The sensitivity of the model with a PHs of 30min or longer had significant differences from the sensitivity of a PH of 15min (\( p < 0.05 \), while the sensitivity with a PHs of 30min, 45min and 60min had no difference (\( p > 0.05 \)). This paper took patient B as an example and its warning result is shown in Figure 6.

**Discussion**

The purpose of blood glucose control is to reach a normal concentration of blood glucose, minimizing the occurrences of hypoglycemia and hyperglycemia, respectively. Hypoglycemia can invoke dangerous situations, is feared by many patients with diabetes and is recognized as a key factor that can lead to failure to reach and maintain good glycemic control, repeated episodes of hypoglycemia can increase the incidence of diabetes–related complications [30]. Therefore, controlling the normal concentration of blood glucose with modern information technology is of great significance for reducing the occurrence of diabetes–related complications.

Based on both the LSTM and GRU models regulating information flow through gate mechanisms, the LSTM–GRU model in this paper was proposed. The results show that a PH of 15min had the highest accuracy and a reduction in prediction accuracy of the three models, as the PH was extended. However, the prediction errors of the three models were all within the acceptable range. A PH of 15min is the most accurate, but they do not provide enough time for clinicians to take action.
when hypoglycemia happened and the models with a PHs of 30min or longer are good enough to allow a patient to do the necessary adjustments in insulin delivery and consequently to prevent the occurrence of hypoglycemia events. Theoretically, the longer PH, the higher the clinical value is. However, due to a reduction in prediction performance over time, the longest time was 60min in this paper.

Correlation analysis can be a good measure to evaluate the degree of correlation between the blood glucose level predicted by the model and the original blood glucose level acquired by CGMS. The results showed the $R$-values of the LSTM-GRU model were identical with different PHs ($R = 0.995$), while the $R$-values of the LSTM and GRU models started to reduce when PH was 30 min or longer, but all values had excellent correlations ($R > 0.5$, $p < 0.001$), which provided certain theoretical support for the LSTM-GRU model to be used for accurate long-term prediction. Zones A and B of the EGA represent the clinically accurate and acceptable zones, respectively. With different PHs, the proportion of the LSTM-GRU model in zones A and B was higher than in the LSTM and GRU models. That is, as the PH was extended, the blood glucose levels predicted by the LSTM-GRU model were closer to the original blood glucose level acquired by CGMS and the prediction effect was the best.

The repeated-measures ANOVA results showed that RMSE values of the LSTM-GRU model had differences from the LSTM and GRU models with different PHs ($p < 0.05$), and RMSE values of the LSTM-GRU model were lower than the LSTM and GRU models with identical PHs. In summary, we can conclude that the LSTM-GRU model had the best prediction performance. The reason might be the structure of the LSTM-GRU model.

### Table 2: Mean prediction performance of the three models with different PHs.

<table>
<thead>
<tr>
<th></th>
<th>15min</th>
<th>30min</th>
<th>45min</th>
<th>60min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>RMSE</td>
<td>MAPE</td>
<td>MAE</td>
</tr>
<tr>
<td>L-G</td>
<td>0.176</td>
<td>0.259</td>
<td>2.652</td>
<td>0.186</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.180</td>
<td>0.267</td>
<td>2.696</td>
<td>0.189</td>
</tr>
<tr>
<td>GRU</td>
<td>0.182</td>
<td>0.270</td>
<td>2.762</td>
<td>0.192</td>
</tr>
</tbody>
</table>

### Table 3: The results of RMSE value variance analysis with different PHs.

<table>
<thead>
<tr>
<th></th>
<th>15min</th>
<th>30min</th>
<th>45min</th>
<th>60min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p$-value</td>
<td>95% CI</td>
<td>$p$-value</td>
<td>95% CI</td>
</tr>
<tr>
<td>1</td>
<td>0.000</td>
<td>(0.005, 0.011)</td>
<td>0.000</td>
<td>(0.005, 0.011)</td>
</tr>
<tr>
<td>2</td>
<td>0.000</td>
<td>(0.007, 0.016)</td>
<td>0.000</td>
<td>(0.006, 0.013)</td>
</tr>
<tr>
<td>3</td>
<td>0.102</td>
<td>(0.000, 0.007)</td>
<td>0.535</td>
<td>(-0.001, 0.004)</td>
</tr>
</tbody>
</table>

### Table 4: The EGA results of the three models (A+B%).

<table>
<thead>
<tr>
<th>Model</th>
<th>15min</th>
<th>30min</th>
<th>45min</th>
<th>60min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>LSTM-GRU</td>
<td>99.71</td>
<td>0.26</td>
<td>99.72</td>
<td>0.25</td>
</tr>
<tr>
<td>LSTM</td>
<td>99.71</td>
<td>0.25</td>
<td>99.71</td>
<td>0.26</td>
</tr>
<tr>
<td>GRU</td>
<td>99.68</td>
<td>0.30</td>
<td>99.68</td>
<td>0.28</td>
</tr>
</tbody>
</table>

### Table 5: Hypoglycemia warning performance of the LSTM-GRU model with different PHs.

<table>
<thead>
<tr>
<th></th>
<th>15min</th>
<th>30min</th>
<th>45min</th>
<th>60min</th>
</tr>
</thead>
<tbody>
<tr>
<td>false-positive rate</td>
<td>0.88%</td>
<td>0.90%</td>
<td>0.87%</td>
<td>0.87%</td>
</tr>
<tr>
<td>false-negative rate</td>
<td>8.79%</td>
<td>10.29%</td>
<td>10.79%</td>
<td>11.27%</td>
</tr>
<tr>
<td>sensitivity</td>
<td>91.21%</td>
<td>89.71%</td>
<td>89.21%</td>
<td>88.73%</td>
</tr>
</tbody>
</table>

### Table 6: Comparison of sensitivity differences of the LSTM-GRU model with different PHs.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>0.016</td>
<td>0.025</td>
<td>0.012</td>
<td>0.533</td>
<td>0.273</td>
<td>0.281</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.291, 2.722)</td>
<td>(0.260, 3.739)</td>
<td>(0.573, 4.394)</td>
<td>(-1.081, 2.068)</td>
<td>(-0.791, 2.746)</td>
<td>(-0.406, 1.374)</td>
</tr>
</tbody>
</table>

Note: A:15min and 30min. B:15min and 45min. C:15min and 60min. D:30min and 45min. E:30min and 60min. F:45min and 60min. 95%CI:95% confidence interval.

The first layer is the LSTM unit, which inherits the learning advantages of the LSTM model in a longer time range. On this basis, The second layer was the GRU unit, which made the sample data easy to train and shortened the operation time. Finally, a fully connected layer (dense layer) was used to connect the hidden layer and the output layer. The LSTM–GRU model made up for the insufficiency of a single model, which improved the overall prediction performance of the model and prolonged the PH.

Based on the efficient prediction performance of the LSTM–GRU model, the hypoglycemia warning was further proposed in this paper. When PH was 15 min, 30 min, 45 min and 60 min, the mean sensitivity of the LSTM–GRU model was 91.21%, 89.71%, 89.21% and 88.73% and Figure 6 confirmed the mean sensitivity of the model with a PHs of 30 min, 45 min and 60 min had no significant difference (p > 0.05). When PH was 30 min or longer, the sensitivity of the hypoglycemia warning of the LSTM–GRU model had subtle changes, which provided the theoretical basis for long-term hypoglycemia warning. The hypoglycemia warning is helpful for clinicians and patients, which provides enough time for clinicians to take action to prevent hypoglycemia events and has excellent clinical application value. The above results may be when PH was 30 min or longer, the sensitivity of the hypoglycemia warning of the model had subtle changes, which resulted in its hypoglycemia warning performance remaining stable with a PHs of 30 min or longer.

Conclusion

The LSTM–GRU model in this paper was proposed for blood glucose prediction and compared its blood glucose prediction performance with the LSTM and GRU models. We found that the LSTM–GRU model performed the best and used the model for hypoglycemia warning, which would help prevent the complications of hypoglycemia and save lives. In future work, we will consider larger datasets, perhaps including many physiological indicators. In addition, we can continue to explore the impact of different hypoglycemia thresholds and blood glucose fluctuations on the experimental results.

Acknowledgments

We appreciate the efforts of every staff member involved.

Author disclosure statement

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.

References


