Smart Continuous Glucose Monitoring Prediction with Grid Long Short-Term Memory

Han Meng
School of Computer Science (National Pilot Software Engineering School), Beijing University of Post and Telecommunications, Beijing, China, 2019211943@bupt.edu.cn, Corresponding author

ABSTRACT

Diabetes is one of the four diseases that are endangering human health, which affects nearly 10% of the world's adults. To solve the continuous glucose monitoring data prediction problem, machine learning methods, including ARIMA, RNN, LSTM, stack LSTM have been applied in CGM prediction. Nevertheless, most of these methods have their deficiency of insufficient accuracy and are sensitive to noise. This paper uses Grid Long Short-Term Memory (LSTM) to predict continuous glucose monitoring data and improve the accuracy obviously. First, the dataset of 30 patients' glucose is preprocessed and is divided into two groups, training set, and test set. To eliminate the negative effect of model architecture complexity, the paper enhances and formats the data to lower the difficulty of the processing part. Then, the proposed Grid LSTM is constructed. Furthermore, the dropout optimization and batching strategy are used to lower lost during training. Finally, the model is tested as well as the lost are calculated and characterized by RMSE and CG-EGA, which is an index of BG level accuracy. To verify the effectiveness of the proposed method, Grid LSTM is compared with ARIMA method. The experimental results show the ARIMA method has a RMSE level of 42.492 whilst the Grid LSTM method has a RMSE level of 33.858, our analysis demonstrates that the proposed method obviously outperforms related baselines, LSTM, RNN, ARIMA, for predicting continuous glucose monitoring.

CCS CONCEPTS

• **Computing methodologies** → Modeling and simulation; Model development and analysis; Modeling methodologies.

KEYWORDS

Grid long short-term memory, continuous glucose monitoring prediction, diabetes, recursive neural networks, blood glucose

ACM Reference Format:
Han Meng. 2022. Smart Continuous Glucose Monitoring Prediction with Grid Long Short-Term Memory. In *7th International Conference on Cyber Security and Information Engineering (ICCSIE2022), September 23–25, 2022, Brisbane, QLD, Australia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3558819.3565076

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICCSIE2022, September 23–25, 2022, Brisbane, QLD, Australia
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9741-4/22/09…$15.00
https://doi.org/10.1145/3558819.3565076

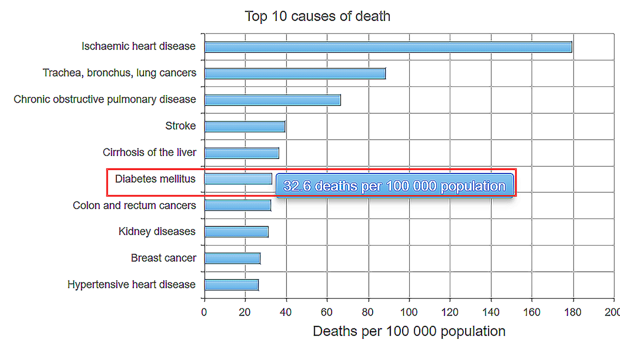

Figure 1: The top 10 death rate in US.

1 INTRODUCTION

Diabetes is one of the four diseases that are endangering human health, and there is no effective method to cure it. The World Health Organization (WHO) estimated that about 300 million people worldwide will have diabetes by 2025 [1]. Diabetes not only brings suffering to the patients and their families but also brings a heavy burden to the country and society. Figure 1 demonstrates diabetes has led to serious damage to our health. To maintain normal blood glucose (BG) levels and reduce complications of diabetes, self-monitoring of BG is necessary for patients with diabetes, therefore BG detection is particularly essential to be real-time, convenient in today's life to improve people's health protection.

Fortunately, powerful new technologies allow for a consistent and reliable treatment plan for people with diabetes. One major development is a system called continuous glucose monitoring (CGM) [2]. However, CGM readings are susceptible to sensor errors [3]. With the development of technology and time, a vast array of research on deep learning (DL) are carried out, which is an essential domain for the development of cutting-edge science and technology. The characteristics learned from large-scale data sets can be well extended to other tasks and data sets, which provides a new opportunity for the BG monitoring and turn BG prediction into possible.

The DL methods include ARIMA [4], RNN [5] and LSTM [6], which motivate the paper to be proposed. ARIMA combines autoregressive model, moving average model and different method to predict based on time series. Furthermore, RNN processes sequence data by recycling neurons, including input units, hidden units and output units. LSTM is an improvement of RNN, which improves the processing performance for the long-term state, that cell state is used to save the long-term state.

In this paper, a new CGM prediction method based on grid LSTM algorithms is proposed. First, 30 groups of CGM data are applied, whilst 23 groups for training models as well as 7 groups for testing.

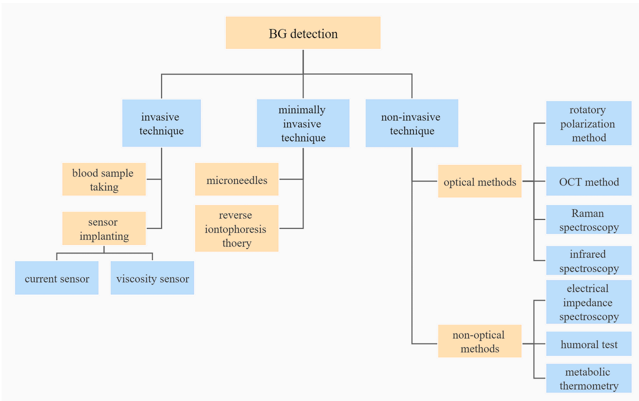

Figure 2: Several methods for BG detection.

As for data preprocessing, Kalman Smoothing to the raw data is useful, and the data is randomly separated to the training group and test group. Second, the model is trained with dropout optimization. Finally, the RMSE, MAE and CG-EGA [7] are calculated in the testing part, which is a standard to measure BG level. Data enhancement and format are applied to reduce the difficulty of preprocessing. To eliminate the adverse effect of high model architecture complexity, the paper enhances and formats the data to lower the difficulty of the processing part. To verify the effectiveness of the proposed method, a comparison is made between ARIMA and the proposed Grid LSTM. Comparing the RMSE level, 33.858 is better than 42.492, which certificates the performance and advance of the Grid LSTM. The experimental results show that this method is significantly better than the relevant benchmark.

2 RELATED WORK

The BG detection methods mainly include invasive technique, minimally invasive technique and non-invasive technique. Moreover, to control the BG level, DL methods are used to monitor blood glucose, which are shown in Figure 2. Furtherly, the ARIMA [4], LSTM [6], ELM [8], GP [9], pcLSTM [10], SVR [11] algorithms are used for CGM prediction. pcLSTM is more efficient and grid LSTM is another variant of LSTM proposed [12], which has not been used in CGM. These models are all the ones that have been verified by MNIST dataset and have high accuracy. Also, the data with smoothing always perform better. The data with smoothing (Kalman Smoothed CGM) and without smoothing (Raw CGM) both passed the test by calculating Area under Curve (AUC), Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), deep RMSE (dRMSE) and Akaike information criterion (AIC), which are reasonable evaluation methods. Moreover, the CG-EGA. The Continuous Glucose Error Grid Analysis (CG-EGA) [7] is also required for reassessment.

3 METHOD

This section introduces the Grid LSTM for Continuous blood glucose monitoring prediction work. Our method mainly contains data preprocessing (Sec. A), training (Sec. B) and model testing (Sec. C).

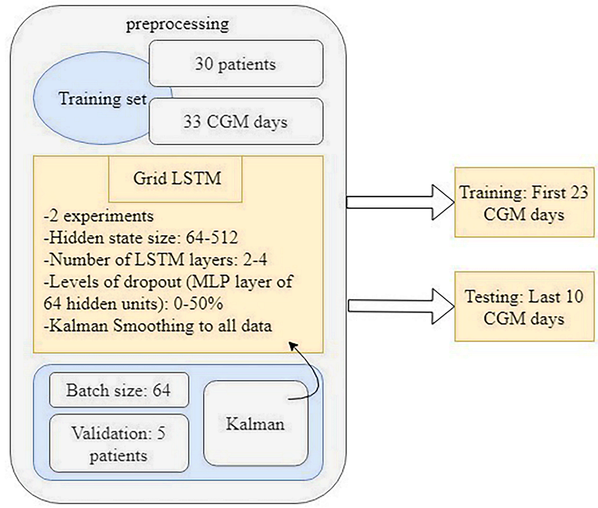

Figure 3: Flow diagram of the experimental setup including (pre)training, validation, testing parts and dataset explanation and instruction.

3.1 Data Preprocessing

The paper includes qualitative analysis and quantitative analysis and uses secondary data from Kaggle. First, we get the general cognition of pieces of data, which involves 30 patients' glucose data and 30 CGM days. The data pieces are divided into 2 separated groups and one piece containing 23 CGM days of 30 patients is applied into training as well as the another piece within 7 CGM days of 30 patients is used for testing. After dividing works, data cleaning, data integration, outlier removal and noise reduction work are conducted in order to find the rules hiding among the dataset. Finally, the data is reordered randomly and forms a new, well-preprocessed dataset. Figure 3 shows the whole process which is arranged during preprocessing.

3.2 Training

This section introduces the training process. Our method mainly contains model architecture and training methods.

1) Model Architecture

As shown in Figure 4 [13], our model mainly includes forget gates, input gates and output gates, and its architecture is described below. Taking positivism methodology as the guidance, the computation and training part is fully based on LSTM units and proceeds as in Equation 1). The model first concatenates the input hidden vectors from the N dimensions:

$$H = \begin{bmatrix} h_1 \\ . \\ . \\ . \\ h_N \end{bmatrix} \tag{1}$$

Then, if a one-dimensional vector is useful for evaluating the entire model, then leave that vector, as well as compute only the outputs of the other dimensions, whilst combining them into one vector.

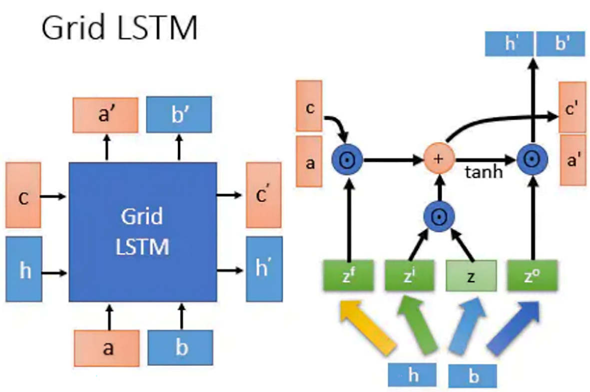

Figure 4: The sketch map of grid LSTM.

Some dimensions do not require *LSTM* training and are passed through the activation function. These nodes are only symbolic connections to the network and have no practical effect. N edges have an input vector and an output vector. Each side of the grid has input or input to it. This mechanism ensures that the hidden layer vectors and the memory vectors of the different edges can be tightly connected without mixing them up.

Furthermore, the block computes N transforms $LSTM\ (\cdot,\ \cdot,\ \cdot)$ [6], one for each dimension, obtaining the desired output hidden and memory vectors [14] as in Equation 2):

$$\begin{bmatrix} (h_1', m_1') = LSTM\,(H, m_1, W_1) \\ \cdot \\ \cdot \\ \cdot \\ \left(h_N', m_N'\right) = LSTM\,(H, m_N, W_N) \end{bmatrix} \tag{2}$$

2) Training Method

First of all, instead of using dropout and regularization items, a small-scale data set is used. The network trains and fits the data set and then checks whether the loss is 0 and the accuracy is 1. In a round of epoch, print out the input and output, detect the correctness of the data, check whether each batch has the same value, and check whether the characteristics correspond to the label. Secondly, visualize the data training. After each round of epoch training, the loss and accuracy of the verification set are calculated, and the loss and evaluation indexes of the training set and verification set after each round of epoch are recorded. Additionally, we adjust the parameters. After ensuring the correctness of the data and the network, the default parameter settings are used to observe the changes of loss, and preliminarily determine the range of each hyperparameter, and then adjust the parameters. For each hyperparameter, we only adjust one parameter each time, and then observe the loss change.

During this process, the following situations may occur. Since the network is overfitted, we use regularization, dropout and batch normalization. We use L2 regularization [15] in this paper. We set dropout to 0.5 on the training set and remove dropout on the verification set and test set. Furthermore, when the network is underfitted, we remove or reduce the regularization, increase the network depth, increase the number of neurons and increase the amount of data in the training set. We use the softsign [16] activation function instead of tanh, and set early stopping to evaluate when stopping early according to the performance on the verification set. We use gradient clipping to normalize the gradient and limit the gradient to 5 or 15.

The grid LSTM model is trained through *pytorch*, whose goal is to construct a network structure using the LSTM hidden layer units and extend depth in any dimension of the network. Like multi-dimensional LSTM, an N dimensional block accepts N hidden layer vectors and N memory vectors as its input. Nevertheless, different from LSTM, the output N hidden layer vectors and memory vectors is explicit. The training process is as follows. Firstly, the output value of each neuron is calculated forwardly. Secondly, the error value of each neuron is calculated in reverse. Then the back propagation of LSTM error term is conducted and it includes two directions: one is the back propagation along time, that is, from the current time t, the error term at each time is calculated. One is to propagate the error term to a forward level. Furtherly, according to the corresponding error term, the gradient of each weight is calculated. Finally, we repeat those steps above and lower the error.

3.3 Model Testing

To verify the effectiveness of the model, we test the model using the following methods. Another piece of data is used in the validation and testing parts. Particularly, in the testing process, we calculate RMSE level and CG-EGA, etc., in order to prove and compute the CGM prediction accuracy. For the testing part, part of BG data is applied to testing. The RMSE is defined as in Equation 3):

$$RMSE\left(x_j^{(s)}, \hat{x}_j^{(s)}\right) = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}\left(x_{i,j}^{(s)} - \hat{x}_{i,j}^{(s)}\right)^2} \tag{3}$$

Finally, the CG-EGA provides a measure of the clinical acceptability of the predictions. Indeed, predictions, depending on the current state of the patient's glycemia (hypoglycemia, euglycemia, or hyperglycemia), can be more or less dangerous, which is a necessary supplement test metric. Eventually, the error is reduced up to the criterion.

4 RESULTS AND DISCUSSION

This section introduces result of our experiments and gives a few discussions. The results and discussion are described as experimental settings (Sec. A) and experimental results (Sec. B).

4.1 Experimental Settings

1) Datasets

This subsection describes the datasets used in our experiment and resources are listed below.

- Diabetes Data Set downloaded from Kaggle originally from the N.Inst. of Diabetes & Diges. & Kidney Dis, containing the daily event and corresponding BG level.
- 6 Months Daily Diabetes Measures downloaded from Kaggle originally from Maxime Fuccellaro, which contains the BG level during morning, noon, and evening.
- Continuous Blood Glucose Monitor Data downloaded from Kaggle originally from Gourav Agrawal, which contains 30 groups of CGM data of 24hours.

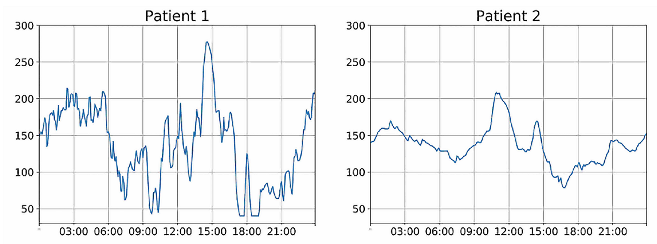

Figure 5: One of the example of data visualization.

Figure 5 demonstrates one of the example of data lists, which shows blood glucose levels of two patients at various time of one day.

2) Baselines

This subsection introduces the baselines used to make comparisons with our method. Grid LSTM is compared with the autoregressive integrated moving average (ARIMA) model. In ARIMA (p, d, q) [4], 'AR' is "autoregressive", and 'p' is the number of autoregressive terms. 'I' is the difference, and 'd' is the number (order) of differences made to make it a stationary sequence. 'MA' is "moving average" and 'q' is the number of moving average items. ACF, the autocorrelation coefficient can determine the value of q, and, PACF, partial autocorrelation coefficient can determine the value of q. ARIMA model is simple, only needs endogenous variables without the help of other exogenous variables, which is easy to implement. It is used as a baseline in this experiment.

Another baseline is LSTM model. Long Short-Term Memory (LSTM) [6], a special recursive neural network (RNN), which is suitable for processing and predicting important events with relatively long intervals and delays in time series.

3) Metrics

Stationary R-square, R-square, RMSE, MAPE, MaxAPE, MAE, MaxAE, Normalization BIC and CG-EGA is used to evaluate and compare the performances of models.

The stationary R-square is more accurate than the R-square if the original sequence is a non-stationary sequence and conforms to the stationary sequence after difference or transformation. R square is applicable to the stationary sequence of the original sequence. R-square is defined as in Equation 4):

$$\mathrm{R-square} = \frac{\mathrm{SSR}}{\mathrm{TSS}} \tag{4}$$

where *SSR* means the sum of squares of residuals, and *TSS* means total sum of squares.

MAPE is defined as in Equation 5):

$$\mathrm{MAPE} = \frac{100\%}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left| \frac{\widehat{\mathrm{y_i}} - \mathrm{y_i}}{\mathrm{y_i}} \right| \tag{5}$$

MAE is defined as in Equation 6):

$$\mathrm{MAE}\,(\mathrm{X}, \mathrm{h}) = \frac{1}{\mathrm{m}} \sum_{\mathrm{i}=1}^{\mathrm{m}} |\mathrm{h}\,(\mathrm{x_i}) - \mathrm{y_i}| \tag{6}$$

Max AE is defined as max absolute error and Max APE is defined as max absolute percentage error.

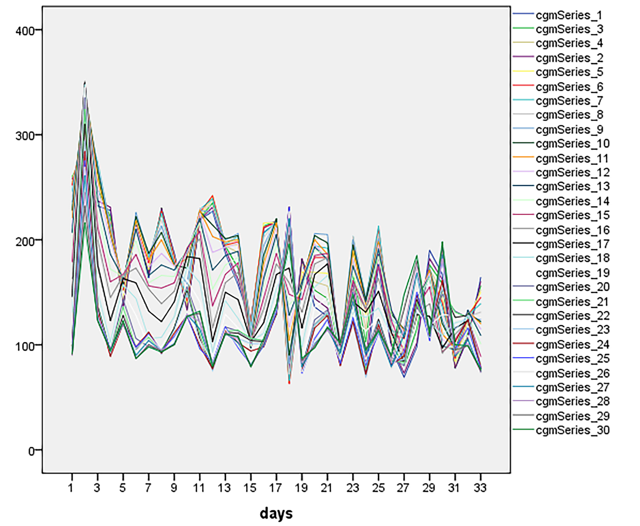

Figure 6: Visualization results of the prediction of 24-33 days blood glucose level

Normalization BIC (Bayesian Information Criterion) is defined as in Equation 7):

$$\mathrm{BIC} = 2\ln(\mathrm{f}(\mathrm{y}|\theta_{\mathrm{k}})) - \mathrm{K}\log\,(\mathrm{n}) \tag{7}$$

Where K means the number of model parameters and n means number of samples.

4.2 Experimental Results

In this paper, 30 people's BG level data is picked up whilst their corresponding data for continuous 33 days is noted down. Furthermore, 23 peoples' blood glucose data for 33 days is applied into training as well as 7 peoples' blood glucose data for 33 days is used in testing work.

Figure 6 shows the results that the curve lies in the first 23 days is the accurate known data and the last 10 days is the prediction value. In order to make a comparison, part of the fitting results, Lower Control Limit (LCL) and Upper Control Limit (UCL) are demonstrated in Figure 7.

Table 1 and Table 2 show the comparison of ARIMA and the Grid LSTM. The mean RMSE is 41.264, which is lower than 42.492 for ARIMA. The MAPE is 10.885 and it's greatly lower than 22.135 for ARIMA. The MAE is 11.441, which is better than 29.839 for ARIMA. To summarize, the Grid LSTM error value and level are lower than its comparison model, ARIMA, which verify the good performance of the proposed model.

In terms of advantages, first of all, the use of knowledge and research methods from multiple disciplines can analyze the subject from different angles, and can better reflect the diversity of influencing factors. In addition, if the observation method is qualitative research, then data mining is quantitative research. The combination of quantitative and qualitative can not only reveal the law but also explore the essence, providing stronger persuasiveness. This model contains multi-dimensional LSTM units and based on previous research results, the test error on the MNIST dataset could

Table 1: Statistic about the model error and evaluation for ARIMA

Fitting Statistics	mean	Standard error	min	max	Percentile						
					5	10	25	50	75	90	95
Stationary R-square	0.489	0.435	-1.146E-13	-908	-9.943E-14	-5.609E-14	6.106E-16	0.640	0.863	0.888	0.902
R-square	0.220	0.200	0.000	0.448	0.000	0.000	-9.88E-5	0.327	0.404	0.442	0.447
RMSE	42.492	3.074	35.972	46.974	36.652	38.098	40.100	42.837	44.825	46.690	46.969
MAPE	22.135	2.750	16.794	27.756	17.180	18.118	20.454	22.150	23.515	26.059	27.459
MaxAPE	89.518	40.361	49.177	178.642	49.738	52.537	58.206	70.083	117.426	165.536	175.130
MAE	29.839	3.597	26.226	36.454	26.320	26.449	27.030	28.218	34.240	35.299	36.265
Max AE	139.37	40.941	85.533	218.970	88.858	95.591	104.737	128.568	170.397	212.326	216.803
Normalization BIC	7.659	.154	7.271	7.825	7.309	7.405	7.557	7.692	7.802	7.815	7.823

Table 2: Statistic about the model error and evaluation for the Grid LSTM

Fitting Statistics	mean	Standard error	min	max	Percentile						
					5	10	25	50	75	90	95
Stationary R-square	0.933	0.030	0.888	0.964	0.888	0.888	0.900	0.936	0.961	0.964	0.964
R-square	0.856	0.050	0.768	0.920	0.768	0.768	0.788	0.858	0.915	0.920	0.920
RMSE	41.264	5.852	33.858	48.576	33.858	33.858	35.140	41.907	48.103	48.576	48.576
MAPE	10.885	1.311	9.166	12.590	9.166	9.166	9.468	10.733	12.072	12.590	12.590
MaxAPE	45.265	3.710	40.231	50.544	40.231	40.231	41.620	44.779	48.084	50.544	50.544
MAE	11.441	1.703	9.170	13.779	9.170	9.170	9.877	10.906	13.184	13.779	13.779
Max AE	37.850	3.829	33.296	42.944	33.296	33.296	33.787	37.986	42.457	42.944	42.944
Normalization BIC	10.455	.286	10.077	10.799	10.077	10.077	10.151	10.503	10.779	10.799	10.799

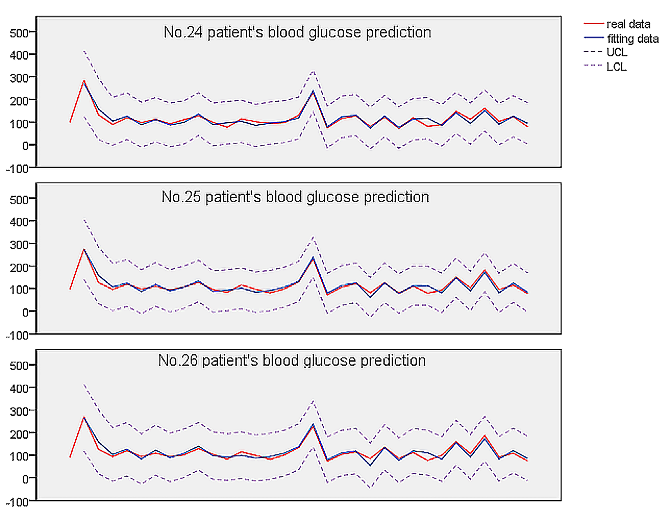

Figure 7: The prediction results for part of test datasets (4 out of 7).

be lowered to 0.28% using dropconnect and with appropriate regularization. Therefore, the performance ensures that will be better than other models, which applied in CGM prediction.

In terms of disadvantages, due to its complexity and data demand, the training process needs a vast array of data with preprocessing, which is difficult to collect. Moreover, the grid LSTM model is used in character prediction, translation and digit image classification domains and gain well-performed results. Nevertheless, the LSTM and pcLSTM applications need to add L2 penalty (10-4) so that the test error is controlled. As mentioned above, compatibility was not proposed before, which is an entirely new topic and represents more challenges.

5 CONCLUSION

In this paper, the continuous blood glucose monitoring data prediction using Grid Long Short-Term Memory is conducted. We propose a new way to predict data about continuous glucose monitoring for those diabetic people, which is conducive to their daily life. Our experimental results show the model's performance is well proved and the usage and intelligence of Grid LSTM algorithm have been well proved within comparison with the other algorithm, ARIMA. Our analyses illustrate that the Grid LSTM model has compatibility to BG prediction.

In the future, our study will focus on the optimization of the model application. Additionally, we will continue to explore the anti-noise capability of the Grid LSTM.

REFERENCES

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

[2] Turksoy, K., Samadi, S., Feng, J., Littlejohn, E., Quinn, L. and Cinar, A., "Meal Detection in Patients With Type 1 Diabetes: A New Module for the Multivariable Adaptive Artificial Pancreas Control System," IEEE Journal of Biomedical and Health Informatics, vol. 20(1), pp. 47-54. 2016.

[3] A. Mohebbi *et al.*, "Short Term Blood Glucose Prediction based on Continuous Glucose Monitoring Data," in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 5140-5145.

[4] De Bois, M., El Yacoubi, M. A., and Ammi, M, "Prediction-coherent lstm-based recurrent neural network for safer glucose predictions in diabetic people," in International conference on neural information processing, Springer, 2020, pp. 510–521.

[5] H. N. Mhaskar, S. V. Pereverzyev, and M. D. Walt, "A function approximation approach to the prediction of blood glucose levels," Frontiers in Applied Mathematics and Statistics, 2021, vol. 7:53.

[6] M. Armandpour, B. Kidd, Y. Du, and J. Z. Huang, "Deep Personalized Glucose Level Forecasting Using Attention-based Recurrent Neural Networks," in 2021 International Joint Conference on Neural Networks (IJCNN), 2021, pp. 1-8.

[7] Nal Kalchbrenner, Ivo Danihelka, and Alex Graves, "Grid long short-term memory," arXiv preprint arXiv:1507.01526, 2015.

[8] Oviedo S, Vehí J, Calm R, and Armengol J. "A review of personalized blood glucose prediction strategies for T1DM patients," Int J Numer Method Biomed Eng, vol. 33(6), 2016.

[9] Ndongo M, Kâ Diongue A, Diop A, and Dossou-Gbété S., "Estimation for seasonal fractional ARIMA with stable innovations via the empirical characteristic function method," Statistics, vol. 50(2), pp.298–311, 2016.

[10] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[11] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," Neurocomputing, vol. 70(1-3), pp. 489–501, 2006.

[12] Edward Pantridge and Lee Spector, "Code Building Genetic Programming," in Proceedings of the 2020 Genetic and Evolutionary Computation Conference (GECCO'20), 2020, pp. 994–1002.

[13] Jiang, J., Yu, L., Zhang, X., Ding, X., and Wu, C, "EV-Based Reconfigurable Smart Grid Management Using Support Vector Regression Learning Technique Machine Learning," Sustainable Cities and Society, vol. 76, 2022.

[14] Glorot, X. and Bengio, Y, "Understanding the difficulty of training deep feedforward neural networks," in Proceedings of 13th International Conference on Artifificial Intelligence and Statistics, 2010, pp. 249–256.

[15] Fang, Y., Wang, Y., and Huang, X, "A divergence formula for regularization methods with an L2 constraint," Mathematics, March 2012.

[16] Zhang, T., and Yu, B. "Boosting with early stopping: Convergence and consistency," Annals of Statistics, vol. 33, pp. 1538-1579, August 2005.