Blood glucose forecasting using LSTM variants under the context of open source artificial pancreas system

Blood Glucose Forecasting using LSTM Variants under the Context of Open Source Artificial Pancreas System

Tianfeng Wang
University of Reading
tianfeng.wang@pgr.reading.ac.uk

Weizi Li
University of Reading
weizi.li@henley.ac.uk

Abstract

High accuracy of blood glucose prediction over the long term is essential for preventative diabetes management. The emerging closed-loop insulin delivery system such as the artificial pancreas system (APS) provides opportunities for improved glycaemic control for patients with type 1 diabetes. Existing blood glucose studies are proven effective only within 30 minutes but the accuracy deteriorates drastically when the prediction horizon increases to 45 minutes and 60 minutes. Deep learning, especially for long short term memory (LSTM) and its variants have recently been applied in various areas to achieve state-of-the-art results in tasks with complex time series data. In this study, we present deep LSTM based models that are capable of forecasting long term blood glucose levels with improved prediction and clinical accuracy. We evaluate our approach using 20 cases(878,000 glucose values) from Open Source Artificial Pancreas System (OpenAPS). On 30-minutes and 45-minutes prediction, our Stacked-LSTM achieved the best performance with Root-Mean-Square-Error (RMSE) marks 11.96 & 15.81 and Clark-Grid-ZoneA marks 0.887 & 0.784. In terms of 60-minutes prediction, our ConvLSTM has the best performance with RMSE = 19.6 and Clark-Grid-ZoneA=0.714. Our models outperform existing methods in both prediction and clinical accuracy. This research can hopefully support patients with type 1 diabetes to better manage their behavior in a more preventative way and can be used in future real APS context.

1. Introduction

Type 1 Diabetes is an autoimmune disease that causes the insulin-producing beta cells in the pancreas to be destroyed, preventing the body from being able to produce enough insulin to adequately regulate blood glucose levels [1]. Estimating and predicting blood glucose in both the short-term and long-term are essential for effective management of diabetes. The traditional approach to managing Type 1 diabetes relies on patients' own estimation of insulin amount which often leads to hyperglycemia or hypoglycemia due to incorrect estimation [2]. The artificial pancreas, or closed-loop insulin delivery system is emerging to continuously monitors blood sugar levels, calculates the amount of insulin required (through a device such as a tablet or mobile phone), and automatically delivers insulin through a pump [3]. Although the insulin pump automatically adjusts basal insulin in existing FDA approved hybrid closed-loop system [4], accurate prediction on long-term blood glucose level under the context of closed-loop artificial pancreas system (APS) is of high importance because it is essential for preventative blood glucose control and to better guide meals intake, exercise and support patients planning daily activities further ahead (e.g. 1 hour). This will allow patients to take actions ahead of time in order to the occurrence of adverse glycaemic events.

Existing blood glucose prediction research focuses on short term predictions such as 15 minutes to 30 minutes but the performance of the prediction models dropped dramatically when it comes to long term predictions such as 45 min to 1 hour [5]. The state-of-the-art deep learning models such as long short term memory (LSTM) and its variants demonstrate strong capabilities in long term forecasting [6]. In this research, we aim to develop a long-term blood glucose forecasting model based on convolutional-LSTM and compare our model with other LSTM models and existing methods used in blood glucose prediction.

URI: https://hdl.handle.net/10125/64139
978-0-9981331-3-3

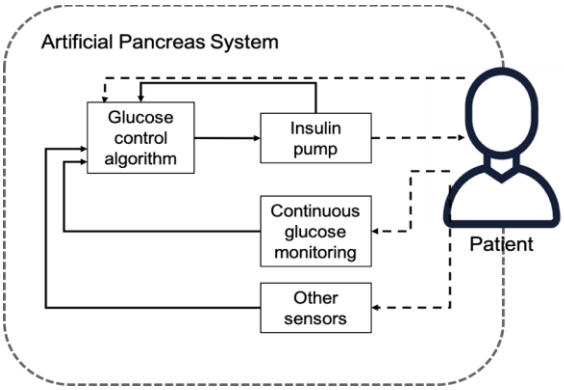

Figure 1 Artificial Pancreas System

The remainder of this article is structured as follows. The related works are presented in Section 2, variations of LSTM are introduced in Section 3, dataset and training process are in Section 4, evaluation methods in Section 5, comments on results in Section 6, and conclusions in Section 7.

2. Blood glucose prediction research

Blood glucose prediction research range from physiological, data-driven and hybrid approach [10]. The physiological approach relies on expert knowledge on insulin and glucose metabolism focusing on simulation models [11, 12, 13]. The main challenge of physiological models is the lack of generalization capability and need support from data for higher prediction performance. Data-driven approaches are mainly based on machine learning methods such as fuzzy logic and rule-based models [14], multi-modal approaches [15, 16] autoregressive models [17, 18], support vector machine [19] and artificial neural networks models [20]. The hybrid approach includes physiological models such as glucose digestion and absorption, insulin absorptions, exercise, and other events. Those physiological models pre-process related data and the results are used in a data-driven model. [21, 22, 23].

Although there are existing studies on blood glucose prediction, the accuracy of longer-term accuracy remains the main challenge for blood glucose prediction studies [4]. Prediction horizon (PH) has been used in the vast majority of the studies for evaluation processes. Existing studies show an increase in the PH leads to a deterioration in the accuracy of a given model [4]. However, PH is important to be considered because patients' needs in deciding meals, physical activity, and other events happen over time. Therefore, both accuracy and PH need to be considered to best meet patients' needs. However, existing research can only demonstrate high performance in 30 min PH but cannot meet the accuracy requirement for glycaemic control for a longer period. Therefore a 30 min PH is the most common value for blood glucose prediction but high accuracy in longer PH is needed.

Deep learning, which incorporates methods recently proved to outperform the already established methodologies [24]. It has led to significant progress in computer vision [25], disease diagnosis [26], and healthcare [27, 28]. Deep learning shows superior performance to traditional ML techniques due to this ability to automatically learn features with higher complexity and representations [29-32]. Recurrent Neural Networks (RNNs) have shown its capability in many applications with time series or sequential data, including machine translation [33, 34] and speech recognition [35]. One of the major challenges in designing systems using classical RNNs is their limited capacity to learn long-term dependencies, because of the vanishing or exploding gradient problem [36]. Recent deep RNNs incorporate mechanisms to address this problem [37], e.g. long-short-term memory (LSTM) which introduces the memory cell and forget gate into classical RNN network [38]. Furthermore, the state-of-the-art LSTM variants such as bidirectional LSTM (Bi-LSTM) [39], vanilla LSTM (V-LSTM) [40], stacked LSTM [41], convolutional LSTM (c-LSTM) [42] and convolutional neural network LSTM (CNN) [43] have shown more promising results for time series predictions [6] because of their capability of capturing rich information from complex time series data. In this research, we propose a deep learning blood glucose prediction model based on LSTM variants for improved prediction and clinical accuracy.

3. LSTM variants based model for long-term blood glucose forecasting

3.1 LSTM

Long short-term memory(LSTM) is a special kind of recurrent neural network architecture(RNN). It is widely used on problems based on time series data such as speech recognition, handwriting recognition, prediction in healthcare pathways, etc. Unlike ordinary RNN, LSTM is specialized at manipulating Long-Term dependencies because it employs the "remember" mechanism through a series of gates. This feature fits the scenario of the glucose prediction problem because the observation window could be quite long which makes other machine learning methods difficult to

handle. The equations for the forward pass of the LSTM unit are as follows.

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$
$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$$
$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$
$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$
$$h_t = o_t \circ \sigma_h(c_t)$$

where the subscript t indexes the time step, the operator$\circ$denotes the element-wise product.

x_t: input vector to the LSTM unit
f_t: forget gate's activation vector
i_t: input/update gate's activation vector
o_t: output gate's activation vector
h_t: hidden state vector also known as output vector of the LSTM unit
c_t: cell state vector
W, U: weight matrices for input vectors and hidden vectors
b: bias vector parameters

The architecture of the vanilla LSTM for glucose prediction is illustrated in Figure 2, a sequence of glucose values are input into the RNN-LSTM network and the target value is predicted at the end of the sequence.

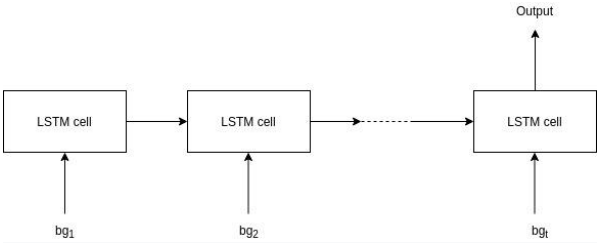

Figure 2 Vanilla LSTM

3.2 Stacked LSTM

The success of deep neural networks attributes to its application of multiple layers. Each layer solves part of the task and altogether the complex network increases the representation power. We can also apply the same strategy on LSTM by adding more layers to make it deeper. The outcome of this idea is the so-called stacked-LSTM. As the name implies, it is an extension of the vanilla LSTM network by stacking a sequence of LSTM layers. Figure 2 gives the architecture of stacked-LSTM, which has several LSTM layers(vertically). In operation, each LSTM layer outputs a sequence of vectors that will be used as the input of the subsequent LSTM layer.

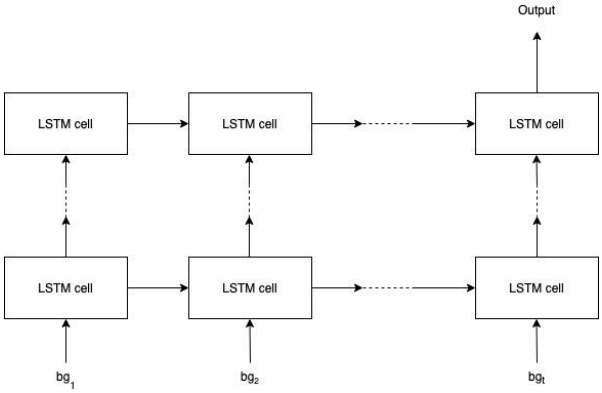

Figure 3 Stacked LSTM

3.3 CNN-LSTM

Due to the intensity of glucose data, we employ CNN in order to better represent the latent features in the glucose series, combined with LSTM we have CNN-LSTM. CNN-LSTM is the combination of CNN layers and LSTM layers in order to take both advantages of CNN and LSTM. It is first designed for spatial inputs prediction problems like image sequence and video sequence prediction, recently it also has been applied in general time series prediction problems and acquired promising results. The architecture of CNN-LSTM as illustrated in Figure 4 includes Convolutional Neural Network(CNN) layers on feature extraction, a follow-up Max Pooling Layer for summarizing the most activated presence of a feature, then a pile of LSTM layers to handle the sequence processing and finally a Fully-Connected Layer before the output.

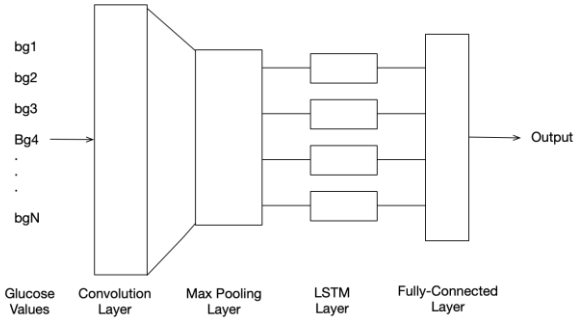

Figure 4 CNN-LSTM

3.4 ConvLSTM

ConvLSTM is another way to leverage both CNN and LSTM. Instead of putting a CNN layer before the LSTM layer, ConvLSTM modifies the internal computation logic and convolution operation in the LSTM cell. In running time, ConvLSTM first read the input with the convolutional part and feed the output into each LSTM unit. The most obvious part exchanged in ConvLSTM is that convolution operations replace matrix multiplication. So we have, e.g. the forget gate becomes $f_t = \sigma_g(W_f * x_t + U_f * h_{t-1} + b_f)$, where '*' denotes convolution. Other formulas listed in Section 3.1 are updated in the same way.
The structure of ConvLSTM presents as follows:

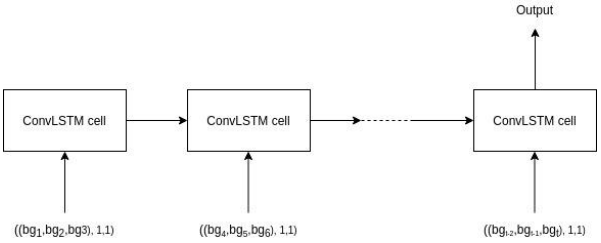

Figure 5 ConvLSTM

The structure looks similar to vanilla LSTM except that the cells are replaced by ConvLSTM cells and the sequence is chunked and shaped to 3 dimensions to meet the needs of convolution calculation.

4. Data and training

The data in this paper comes from donated CGM data of the OpenAPS project [7,8,9]. The glucose values are recorded every 5 minutes. We selected twenty persons' datasets from youth and adult age groups respectively which have the most integrity in one period. The overall number of time points is 878k, of which 1/3 are reserved as test data. Regarding the setting of the training and target window, we consulted clinicians about their practice in evaluating patients' glucose history and make a decision as follows: The training sliding window sizes are 60 minutes and 120 minutes respectively as they demonstrated better performance than other window sizes. So the input represented with time steps is 12 and 24. The prediction horizons are 30 minutes, 45 minutes and 60 minutes respectively. That gives the output length is 6, 9 and 12. There are limited missing values in the datasets, we filled them with linear interpolation

The hardware for the training task includes 1 x NVIDIA Tesla V100 and 4vCPU 26G memory, e.g.

The software that we used includes Pandas for data wrangling and Keras-LSTM library for training, the batch size is 128, the number of epochs is 100.

5 Evaluation

We evaluate the results from two perspectives. One is the statistical evaluation which we use root-mean-squared-error(RMSE) to evaluate the prediction ability of the model. Another is the clinical accuracy evaluation which we employ the Clarke error analysis.

RMSE is the square root of the average squared difference between predicted values and the actual values. In general, the lower this value means a better average prediction performance. The RMSE formation can be illustrated as follows:

$$RMSE = \sqrt{1/n \sum_{k=1}^{n} (y_k - \hat{y}_k)^2}$$

where y_k is the actual value and $\hat{y}_k$ is the predicted one. Although RMSE is widely used in the evaluation of time series prediction, it takes each value equally and only looks at the value difference. However, in medical practice like glucose management, different values may have significant difference in clinician outcome. Thus we introduce Clarke Error Grid analysis which pays more attention to the medical significance and amplify the prediction errors that could lead to risk treatments. As shown in figure 6, 7 and 8 of the error grids, the horizontal axis represents true blood glucose and the vertical axis represents predicted blood glucose by the model. Specifically, breaks down the true-predicted blood glucose value scatter plots into five clinical meaningful regions. The regions signify the degree of risk posed by the incorrect prediction.

- Section A. Predicted blood glucose value is within 20% of the actual blood glucose values. This means the prediction error has no effect on clinical action therefore these points are also called clinical accurate ones which are appropriate to lead to the interventions.
- Section B. Predicted value is beyond 20% but would not lead to inappropriate treatment. The prediction error has little or no effect on clinical outcomes.
- Section C. The points in this area indicate the prediction errors might indicate an unnecessary treatment.
- Section D. The points in this area means the prediction errors will lead to a dangerous failure of detecting hypoglycemia and hyperglycemia

● Section E. The points in this area means the prediction error could lead to dangerous consequences and it will confuse the treatment of hypoglycemia and hyperglycemia.

Of all the 5 sections, the percentage of A+B states how the prediction algorithm performs in a clinical acceptance way, while we should also be aware of C, D, and E which symbolizes the errors that may lead to miss judgment in treatments. More percentage in A means less errors thus more clinically accurate predictions.

6. Results

6.1 Prediction accuracy in RMSE

We compared the performance among LSTM variants based models and support vector regression (SVR) method as the baseline. Table 1 shows the blood glucose prediction accuracy of CNN, v-LSTM, CNN-LSTM convLSTM and SVR over the prediction horizon of 30, 45 and 60 minutes. We notice that stacked-LSTM gives the best RMSE performance under the short-term(30 minutes) and mid-term(45 minutes) horizon. Vanilla-LSTM has the lowest score on long-term(60 minutes) horizon prediction. However, vanilla-LSTM only outperforms stacked-LSTM with 0.1 difference(19.01 vs 19.24) .If we consider the overall performance on all prediction horizons, stacked-LSTM achieves the best score. Besides, all the LSTM variant models outperform SVR in 30, 45 and 60 prediction horizons.

Table 1 Blood Glucose Prediction Accuracy comparison in RMSE

Method\PH(min)	30 min	45 min	60 min
CNN	14.74±1.06	18.08±1.94	21.04±2.45
vanilla-LSTM	12.33±1.15	15.86±1.80	**19.01**±2.62
stacked-LSTM	**11.96**±1.02	**15.81**±1.56	19.24±1.78
CNN-LSTM	13.05±1.21	16.72±2.28	19.80±2.54
convLSTM	12.20±0.94	15.82±1.85	19.60±2.01
SVR	13.28±1.02	17.89±1.34	24.21±2.96

6.2 Clinical Accuracy

In addition to prediction accuracy, we also evaluated the clinical accuracy using the Clarke Error analysis to understand the clinical value of the proposed methods. Table 2 shows the score of Clarke zone A and zone B of CNN, v-LSTM, bi-LSTM, s-LSTM, CNN-LSTM, convLSTM and SVR over 30 minutes, 45 minutes and 60 minutes prediction horizon.

Table 2 The Clinical Accuracy from Clarke Error Analysis

Method\ PH(min)	30 min		45 min		60 min	
	Azone	Bzone	Azone	Bzone	Azone	Bzone
CNN	0.844	0.124	0.732	0.230	0.652	0.304
vanilla-LSTM	0.871	0.108	0.782	0.182	0.650	0.309
stacked-LSTM	**0.887**	0.089	**0.784**	0.181	0.700	0.250
CNN-LSTM	0.861	0.110	0.748	0.214	0.700	0.257
convLSTM	0.868	0.102	0.782	0.182	**0.713**	0.245
SVR	0.804	0.107	0.703	0.180	0.645	0.213

We can learn that all methods have a high clinical acceptance rate because the total score of zone A and B for each LSTM variant based model has an average above 0.95. When we look at the long term prediction accuracy, convLSTM gives the best performance with zone A score of 0.713 in 60 minutes prediction horizon. Stacked-LSTM shows the best performance in 30 and 45 minutes prediction horizons with Clarke zone A score of 0.887 and 0.784 respectively. Figure 6-8 illustrates points distribution of the best prediction on Clarke Grid Analysis on 60 minutes, 45 minutes, and 30 minutes prediction horizons respectively. It's clear that the majority of points spread in ZoneA+B which is good to lead the treatment. When we look at Zone C-D, compared with zone-C, points zone-D develops fast as the target horizon increases from 30 minutes to 60 minutes. It indicates that the failure of detecting hypoglycemia and hyperglycemia increases as we predict farther in the future. That is one potential orientation for optimization.

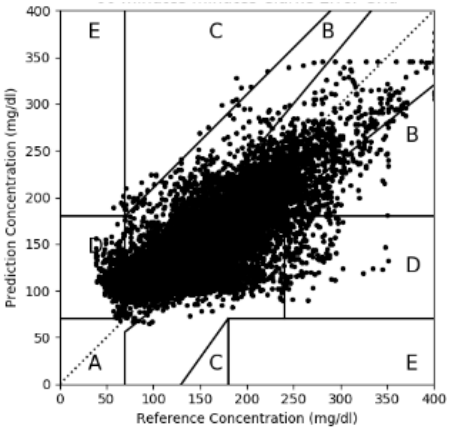

Figure 6 Clarke Error of convLSTM on PH 60(min)

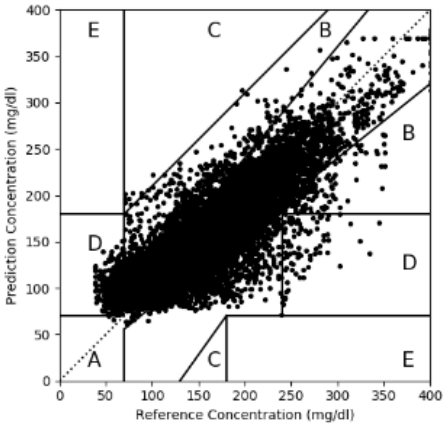

Figure 7 Clarke Error of Stacked LSTM on PH 45(min)

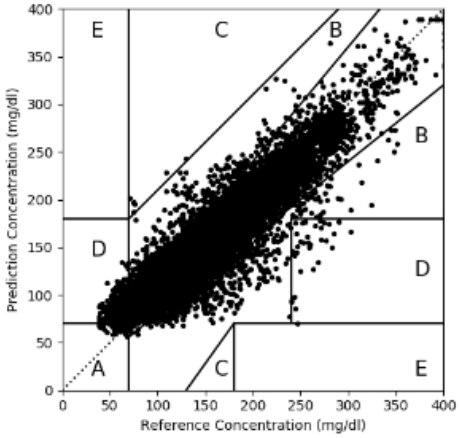

Figure 8 Clarke Error of Stacked LSTM on PH 30(min)

7. Conclusion

In this paper, we developed LSTM variants based blood glucose prediction models for improved prediction and clinical accuracy in long prediction horizon. The modified LSTM can capture more long-term dependencies due to deeper architecture and learn to remove background noise, outline important features and better captures both future and past context of the input sequence. We evaluate the prediction and clinical accuracy of the long term (above 30 minutes) of the proposed methods using 20 cases of real-life data from OpenAPS community. Prediction results were compared between LSTM variants and to those established learning algorithms and widely used algorithms applied to the real-time prediction of glucose using CGM data. Prediction Horizons (PH) of 30, 45, and 60 minutes were used. The proposed LSTM variant based methods showed superior performance in forecasting BG levels (RMSE and clinical accuracy) against existing methods. For several other works, it is difficult to evaluate the RMSE through direct comparison due to the availability of benchmark datasets. However, we may compare the results with widely used methods as benchmarks, such as SVR. The results show that our approach suggests superiority in their prediction accuracy over the 30, 45, and 60 minute time period than existing studies [43][46]. As far as we know, the proposed algorithm achieves a performance state-of-the-art accuracy with regard to RMSE and clinical accuracy.

There are several limitations and future work for this research. First, the longest prediction horizon evaluated is 60 minutes and future work will further improve the proposed models for longer term prediction towards more than 4 hours. Second, future life events will be considered over the longer prediction horizons to improve the performance. A hybrid model combining the advantages of both physiological and LSTM based approach could be developed. Thirdly, there is timing effects that users of OpenAPS have DIY systems making insulin dosing adjustments and acting upon them so it will affect the prediction results. We will consider quantify these influences and make the prediction more accurate. Although some works suggest that ingested carbohydrate information, along with injected insulin information might be redundant [44, 45], we will in future incorporate more clinical information such as comorbidities and other information from Electronic Patient Record for detailed patients

phenotyping and personalized prediction model development. Finally, we have demonstrated the application of deep learning based blood glucose prediction model in the real-life data but more data from both OpenAPS and patients under various closed-loop system could reflect a wider population.

9. References

[1]Diabetes UK. Type 1 diabetes. https://www.diabetes.co.uk/type1-diabetes.html. Accessed 14 June 2019.

[2] Blauw H, Keith-Hynes P, Koops R, DeVries JH. A review of safety and design requirements of the artificial pancreas. Annals of biomedical engineering. 2016 Nov 1;44(11):3158-72.

[3] Diabetes UK. Research spotlight – the artificial pancreas. https://www.diabetes.org.uk/research/research-round-up/research-spotlight/research-spotlight-the-artificial-pancreas. Accessed 14 June 2019.

[4] Medtronic MiniMed Inc. Summary of Safety and Effectiveness Data MiniMed 670G System.; 2016.

[5] Oviedo S, Vehi J, Calm R, Armengol J. A Review Of Personalized Blood Glucose Prediction Strategies For T1dm Patients. Int j numer method biomed eng. 2016; pmid:27644067

[6] Salehinejad H, Sankar S, Barfett J, Colak E, Valaee S. Recent advances in recurrent neural networks. arXiv preprint arXiv:1801.01078. 2017 Dec 29.

[7] OpenAPS. Open APS. 2017. Available at: https://openaps.org/ Accessed February 28, 2018.

[8] Litchman ML, Lewis D, Kelly LA, Gee PM. Twitter analysis of# OpenAPS DIY artificial pancreas technology use suggests improved A1C and quality of life. Journal of diabetes science and technology. 2019 Mar;13(2):164-70.

[9] Lewis D. OpenAPS Outcomes. OpenAPS.org. 2018.

[10] Contreras I, Oviedo S, Vettoretti M, Visentin R, Vehí J. Personalized blood glucose prediction: A hybrid approach using grammatical evolution and physiological models. PloS one. 2017 Nov 7;12(11):e0187754.

[11] Man CD, Micheletto F, Lv D, Breton M, Kovatchev B, Cobelli C. The UVA/PADOVA Type 1 Diabetes Simulator: New Features. J Diabetes Sci Technol. 2014;8: 26–34. pmid:24876534

[12] Fernandez M, Villasana M, Streja D. Glucose dynamics in Type I diabetes: Insights from the classic and linear minimal models. Comput Biol Med. 2007;37: 611–627. pmid:16867301

[13] Ghosh S. A differential evolution based approach for estimating minimal model parameters from IVGTT data. Comput Biol Med. 2014;46: 51–60. pmid:24529205

[14] Fong S, Mohammed S, Fiaidhi J, Kwoh CK. Using causality modeling and Fuzzy Lattice Reasoning algorithm for predicting blood glucose. Expert Syst Appl. 2013;40: 7354–7366.

[15].Buckingham B, Chase HP, Dassau E, Cobry E, Clinton P, Gage V, et al. Prevention of Nocturnal Hypoglycemia. Diabetes Care. 2010;33: 1013–1017. pmid:20200307

[16]Wang Y, Wu X, Mo X. A novel adaptive-weighted-average framework for blood glucose prediction. Diabetes Technol Ther. 2013;15: 792–801. pmid:23883406

[17].Lu Y, Gribok A V., Ward WK, Reifman J. The importance of different frequency bands in predicting subcutaneous glucose concentration in type 1 diabetic patients. IEEE Trans Biomed Eng. 2010;57: 1839–1846. pmid:20403780

[18]Novara C, Pour NM, Vincent T, Grassi G. A Nonlinear Blind Identification Approach to Modeling of Diabetic Patients. Proc 19th World Congr Int Fed Autom Control. 2015; 1–9.

[19] Clarke, William L. "The original Clarke error grid analysis (EGA)." Diabetes technology & therapeutics 7.5 (2005): 776-779.

[20].Fernandez de Canete J, Gonzalez-Perez S, Ramos-Diaz JC. Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes. Comput Methods Programs Biomed. 2012;106: 55–66. pmid:22178070

[21]Balakrishnan NP, Rangaiah GP, Samavedham L. Personalized blood glucose models for exercise, meal and insulin interventions in type 1 diabetic children. 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2012. pp. 1250–1253. 10.1109/EMBC.2012.6346164

[22] Estrada GC, Kirchsteiger H, Eric R. Innovative Approach for Online Prediction of Blood Glucose Profile in Type 1 Diabetes Patients. Am Control Conf (ACC), 2010. 2010; 2015–2020.

[23] Zecchin C, Facchinetti A, Sparacino G, Cobelli C. Jump neural network for online short-time prediction of blood glucose from continuous monitoring sensors and meal information. Comput Methods Programs Biomed. Elsevier Ireland Ltd; 2014;113: 144–152. pmid:24192453

[24] Li K, Daniels J, Liu C, Herrero-Vinas P, Georgiou P. Convolutional recurrent neural networks for glucose prediction. IEEE Journal of Biomedical and Health Informatics. 2019 Apr 1.

[25] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in Proceedings of the 22nd ACM International Conference on Multimedia, ser. MM ’14. New York, NY, USA: ACM, 2014, pp. 675–678.

[26] G. Litjens, C. I. Sanchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. H. van de Kaa, P. Bult, B. van Ginneken, and J. van der Laak, “Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis,” Scientific Reports, vol. 6, p. 26286, May 2016.

[27] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, “Deep learning for healthcare: review, opportunities and challenges,” Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236–1246, 05 2017.

[28] T. Zhu, K. Li, P. Herrero, J. Chen, and P. Georgiou, “A deep learning algorithm for personalized blood glucose prediction,” in The 3rd International Workshop on Knowledge

Discovery in Healthcare Data, IJCAI-ECAI 2018, Stockholm, Sweden, July 2018.
[29] Y. Bengio, "Deep learning of representations: Looking forward," in Statistical Language and Speech Processing. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 1–37.
[30] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85 – 117, 2015.
[31] K. Li, A. Javer, E. Keaveny, and A. Brown, "Recurrent neural networks with interpretable cells predict and classify worm behaviour," in Workshop on Worm's Neural Information Processing (WNIP) in NIPS, CA, USA, 2017.
[32] Q. Zhang, Y. N. Wu, and S.-C. Zhu, "Interpretable convolutional neural networks," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2018, pp. 8827–8836.
[33] Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. ICLR (2015).
[34] Sutskever, I., Vinyals, O. & Le, Q. V. Sequence to sequence learning with neural networks. In Advances in neural information processing systems, 3104–3112 (2014).
[35] Hinton, G. et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Process. Mag. IEEE 29, 82–97 (2012).
[36] Y. Bengio, P. Simard, and P. Frasconi, "Learning Long-Term Dependencies with Graident Descent is Difficult," Saudi Med J, vol. 33, pp. 3–8, 2012.
[37] S. Park, S. Min, H.-S. Choi, and S. Yoon, "Deep Recurrent Neural Network-Based Identification of Precursor microRNAs," Nips, no. Nips, 2017.
[38] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
[39] P. Su, X. Ding, Y. Zhang, F. Miao, and N. Zhao, "Learning to Predict Blood Pressure with Deep Bidirectional LSTM Network," pp. 1–19, 2017.
[40] Wu Y, Yuan M, Dong S, Lin L, Liu Y. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. Neurocomputing. 2018 Jan 31;275:167-79.
[41] A. Graves, "Generating sequences with recurrent neural networks," arXiv preprint arXiv:1308.0850, 2013.
[42] G. Trigeorgis, F. Ringeval, R. Brueckner, E. Marchi, M. A. Nicolaou, B. Schuller, and S. Zafeiriou, "Adieu features? end-to-end speech emotion recognition using a deep convolutional recurrent network," in Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE, 2016, pp. 5200–5204.
[43] Sainath TN, Vinyals O, Senior A, Sak H. Convolutional, long short-term memory, fully connected deep neural networks. In2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015 Apr 19 (pp. 4580-4584). IEEE.
[44] E. I. Georga, V. C. Protopappas, D. Ardigo, M. Marina, I. Zavaroni, D. Polyzos, and D. I. Fotiadis, "Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression," IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 1, pp. 71–81, Jan 2013.
[45] Zecchin, C Online glucose prediction in type 1 diabetes by neural network models. Univ. Degli Stud. Di Padova. Sch. Inf. Eng. Sect. Bioeng. XXVI Ser.; 2014.
[46] Lu, Y, Gribok, AV, Ward, WK, Reifman, J. The importance of different frequency bands in predicting subcutaneous glucose concentration in type 1 diabetic patients. IEEE Trans Biomed Eng. 2010; 57(8): 1839– 1846.
[47] K. Plis, R. Bunescu, C. Marling, J. Shubrook, and F. Schwartz, "A machine learning approach to predicting blood glucose levels for diabetes management," in Modern Artificial Intelligence for Health Analytics Papers from the AAAI-14.