Realizing Blood Glucose Prediction by Convolutional Recurrent Neural Networks with Residual Blocks

Yuxi Zhang

***Abstract*—An advanced convolutional recurrent neural network architecture for forecasting blood glucose is proposed in this paper. To improve the competitiveness of the suggested model, several merits of WaveNet, a deep learning model that performs well on processing audio waveforms, are adapted and implemented. To be more specific, a multi-layer dilated causal convolutional neural network (CNN) with residual blocks followed by a modified recurrent neural network (RNN) with GRU cells is the architecture of our model. 10 virtual adult patients from the UVA/Padova T1D simulator provide the 10 simulated datasets for in-silico experiment, and each dataset consists of 6 channels of time series data, including glucose levels, insulin dosages, carbohydrate intake, the rate of glucose appearance, plasma insulin and the time index. After preprocessing, the data is fed into the network with the aim to forecast the blood glucose level in the next 30 minutes. The obtained prediction results are evaluated by the root mean squared error (RMSE). The average of the best RMSE among the 10 subjects is 8.3050. This RMSE result is better than that of many current algorithms using the same datasets, which shows the superior performance of ourproposed model.**

***Index Terms*—dilated causal CNN, residual learning, gatedrecurrent unit (GRU), glucose prediction**

I. Introduction

Blood glucose (BG) has historically been an essential indicator of diagnosis of diabetes, and for patients with Type 1 diabetes which is widely thought to be precipitated by the destruction of pancreatic β cell that is responsible for producing insulin, lifelong glucose management is required [1]. With the rapid development of Artificial Intelligence, machine learning algorithms such as support vector regression [2] and artificial neural networks (ANN) [3], [4] have been used to help predict glucose for patients with Type 1 diabetes. However, the performance of support vector regression degrades when handling a large training dataset. And because there are typically less than 3 layers in ANN according to literature, it is hard to process a large and complex dataset as well. To enhance the model complexity and hence improve the ability to representation, deep learning networks, composed of multiple hidden layers and neurons, have been explored recently and shown outstanding performance [5]-[8].

In this work, a novel deep learning model for glucose prediction is developed. This model is designed by incorporating the merits of both a convolutional recurrent neural network (CRNN) and a deep neural network (DNN) model called WaveNet. The CRNN architecture is proposed to solve BG prediction problems in 2019 [9] and the DNN model is firstly introduced by the company DeepMind to generate raw audio waveforms [10]. The input datasets are collected from 10 virtual adult patients with diabetes, and each of them is composed of 6 fields, which are glucose levels, insulin dosages, meals intake, the rate of glucose appearance, plasma insulin, and the time index. After preprocessing, the data is fed into the proposed model for training with the aim to predict the BG level in the next 30 minutes. There are two parts in the architecture of this modified CRNN: a 20-layer dilated causal CNN with residual blocks to capture patterns, followed by an RNN including Gated Recurrent Unit (GRU) cells and fully connected layers. Those residual blocks can effectively solve the degradation problem of accuracy in a deep-layer architecture [11].

As far as we know, it is the first time that residual blocks have been added into a CRNN architecture to realize glucose prediction. And our method shows competitiveness when comparing the RMSE with that of another three machine learning algorithms, which are support vector regression (SVR) [2], neural network predictive glucose (NNPG) [3], and bidirectional recurrent neural network [12]. The result section illustrates the superior performance of the proposed model in detail.

Manuscript received October 9, 2021; revised May 11, 2022.

Yuxi Zhang is with the Xi'an Jiaotong-Liverpool University, China (e-mail: Y.Zhang456@student.liverpool.ac.uk).

II. Database and Pre-processing

A. Database

The data used for this paper is generated from the UVA/Padova T1D simulator, which is the only emulator for glucose level simulation approved by the Food and Drug Administration (FDA). 10 virtual adult subjects are included to pretend patients with Type 1 diabetes. We created data for 360 days with three meals per day. And the data is sampled every 5 minutes, therefore, 288 time-series points are collected per day. The virtual glucose dataset is composed of 6 channels of inputs: the glucose levels, the insulin dosages, carbohydrate intake, the time index and another two fields—the rate of glucose appearance and plasma insulin. The plasma insulin controls the production of glucose, and the rate of glucose appearance is a direct indicator for BG prediction [13]. It is noticed that the insulin entries vary from 1 to 5 in each day with the meal, which suggests that meal and insulin injection are given at almost the same time. Besides, the target labels in our model are the changes of the current BG level $x(t)$ and the BG level after 30 minutes $x(t+6)$. The temporary output of the proposed model is the change of BG level Δx between $x(t)$ and $x(t+6)$, therefore, the real prediction of BG level at time $t+6$ can be obtained as $x(t+6) = x(t) + \Delta x(t)$.

doi: 10.18178/ijmlc.2022.12.6.1114

B. Filtering and Normalization

Before being fed into the CRNN model, the time series data experiences some preprocessing. A median filtering is applied to the glucose data to remove outliers in the dataset. And normalization is used to constrict large values within (0,1) to avoid that the prediction results are determined significantlyby these large values.

III. Methodology

CRNN architecture is such a piece of pioneering work in glucose prediction that was firstly proposed in 2019 [9]. It combines the strengths of both CNN and RNN and primarily consists of two modules: a multi-layer CNN, and an RNN block with long short-term memory (LSTM) cells and fully connected layers. Our suggested model retains the basic structure of the original CRNN while modifying the specific algorithms applied in both CNN and RNN parts after being enlightened by WaveNet. The changes in the novel CRNN model can be summarized into three aspects: extracting patterns using dilated causal convolutional layers, solving the degradation problem of deep CNN using residual blocks, and forecasting time series using GRU. These approaches are explained in detail as following and the complete model structure is demonstrated at the end of this section.

A. Dilated Causal Convolutional layers

Convolutional neural networks (CNN) as essential algorithms of deep learning, witness promising advantages of detecting time series data features [14]. Instead of the conventional convolutional neural network, dilated causal CNN is employed in this work. Causal convolutional layers are the main ingredients in WaveNet [10]. By using a causal convolutional network, a model with time-series data such as the proposed model can acquire predictions from a few correlated inputs instead of solely one input.

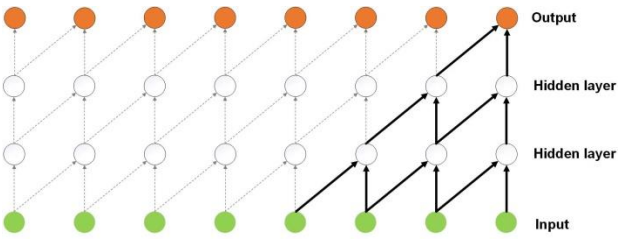

Fig. 1. Causal convolutional neural network with 3 layers [10].

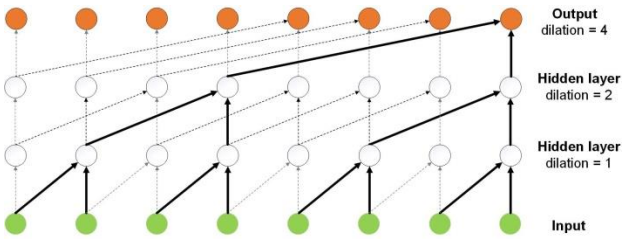

Fig. 2. Dilated causal CNN with 3 layers [10].

A dilation convolution, which is also called convolution with holes, refers to a convolution with a larger filter obtained by dilating the original filter with zeros [10]. By adding the dilated convolutions into the causal convolutional layers, networks with larger receptive fields can be achieved without significantly increasing the computation. To be more specific, as shown in Fig. 1 and Fig. 2, the receptive field increases to 8 after including dilations at various layers while it is only 4 (= #layers + kernel size - 1) before including dilations.

In this work, motivated by the idea of involving more previous inputs to enhance the efficiency and accuracy of the predictive model, 1, 2 and 4 dilations are respectively employed into the hidden layers within each residual block (refers to Fig. 7). And after being processed by the dilated causal convolutional layers, the filtered input signals are transformed into vectors with certain features for further being fed into the followed recurrent layers.

B. Residual Blocks

The depth of neural networks can greatly influence their performance. Intuitively, the deeper the network, the more accurate the model can be since more complex pattern extractions are performed. However, it is found that excessive layers added into the network can cause the accuracy to degrade promptly when it has already become saturated. To solve this degradation problem, residual blocks are proposed by applying the idea of "shortcut connection", which can be easily interpreted as skipping several convolutional layers in the network and directly adding the inputs into the outputs of the stacked layers [11]. The structure on the left-hand side of Fig. 3 represents a building block of residual learning with a shortcut connection. This residual configuration eases the learning process and makes deep neural networks are relatively simple to train.

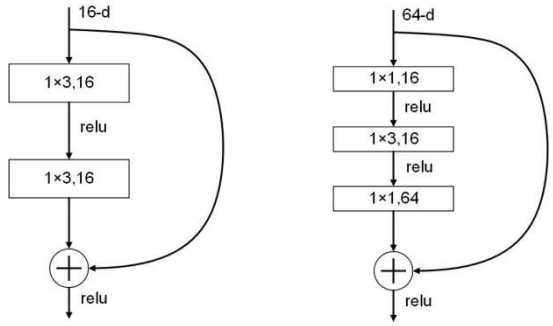

Fig. 3. Two different configurations of a residual block.

In the proposed model, there are 20 one-dimensional convolutional layers including 6 residual blocks (6 shortcut connections). The details of our CNN are shown in Appendix (Fig. 7). It is noticed that a modified block with three layers is used instead of the original two. Fig. 3 compares the two different configurations. The three layers witness different filter length, which is 1×1, 3×3, and 1×1 respectively, and the output dimensions at first two layers are reduced to 1/4 of the original while at the last layer is recovered to the original. This bottleneck design enables the CNN to extract enough features while decreasing the computational cost. By our experiment, it is verified that the suggested model with these residual blocks outperforms the one without them: the RMSE of the former is 8.31 whereas of the latter is 18.56. The evaluation indicator RMSE is defined explicitly in the results section.

C. A Recurrent Layer with GRU Cells

Recurrent neural networks (RNN) are good at processing and predicting time series data. By passing the partial

resultant at time t to the next time t+1, the internal correlation of the time series can be considered while forecasting. For example, a certain word can refer to different meanings due to different contexts. Therefore, to obtain the correct information in time-series data, RNN should be employed.

To achieve better information dissemination, Gated Recurrent Unit (GRU) as one of the variants of RNN is applied in our work. The structure of GRU [15] is shown in Fig. 4.

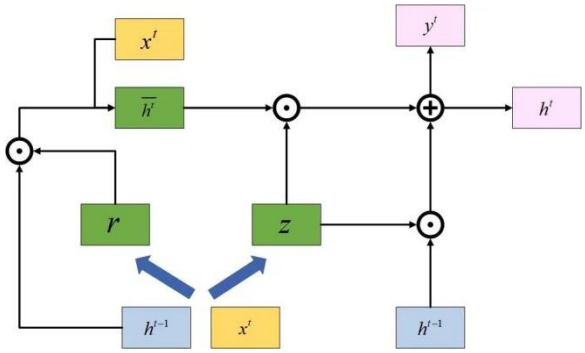

Fig. 4. Structure of Gated Recurrent Unit (GRU).

In the diagram, h^{t-1} represents the data passed from the last layer while h^t represents the data passes to the next layer. x^t is the input and y^t is the output of this layer. r refers to the reset gate, and z refers to the update gate. Both gates receive x^t and h^{t-1} to be the inputs, and the mathematical forms can be written as

$$r^t = \sigma(W^r x^t + U^r h^{t-1}) \quad (1)$$

$$z^t = \sigma(W^z x^t + U^z h^{t-1}) \quad (2)$$

$$\overline{h^t} = tanh(W x^t + U\,(r^t \times h^{t-1})) \quad (3)$$

where σ is the sigmoid activation function, and W and U are the corresponding optimal weights through training. $\overline{h^t}$ stands for a candidate activation that acts as an intermediate variable, and it receives x^t, h^{t-1} and r^t as the inputs. *tanh* is the tanh activation function. X is the matrix multiplication. The final output can be expressed as

$$h^t = (1 - z^t) \times h^{t-1} + z^t \times \overline{h^t} \quad (4)$$

The form above implies the update gate can perform remembering and forgetting information at the same time, and the remembering and forgetting process at each layer can be controlled more flexibly. This also explains the reason why GRU is adopted in the proposed model instead of LSTM which is also a recurrent layer that has been widely used: the calculation cost of GRU is cheaper than that of LSTM since the number of gates in GRU is less than LSTM (3 gates included); and because of the reset gate, its actual accuracy can be higher than LSTM [16].

In this work, it is manifested that the recurrent network with 64 GRU cells experiences a faster speed, and the result of RMSE of it surpasses the one with 64 LSTM cells. The discussion section shows the comparison explicitly.

D. System Architecture

The complete architecture of the proposed predictive model is shown in Fig. 5. The data at the left in the diagram is concatenated into time series data of 6 fields, which are the glucose levels, the insulin dosages, carbohydrate intake, the time index, the rate of glucose appearance and plasma insulin. After preprocessing (filtering and normalization), these data are fed into the modified multi-layer CNN. The details such as the dimension of data in each layer in the CNN have been discussed and shown in Fig. 7. Then the time-series resultant of CNN with the concatenation of features is sent to the modified RNN, which consists of 64 GRU cells and three fully connected layers. After processing by RNN, the predictive glucose level in predictive horizon (PH) = 30 minutes is obtained.

To be more specific, a max pooling is utilized after every three residual blocks of CNN to remove trivial but remain important features to decrease the computational cost. And a dropout is applied after the GRU layer and after the first two fully connected layers to avoid over-fitting issues by randomly dropping units from the neural network during the training process [17]. Besides, RMSprop as an optimizer that is widely applied to deep×learning is used in the proposed model [18]. The learning rate of it is set to 0.002.

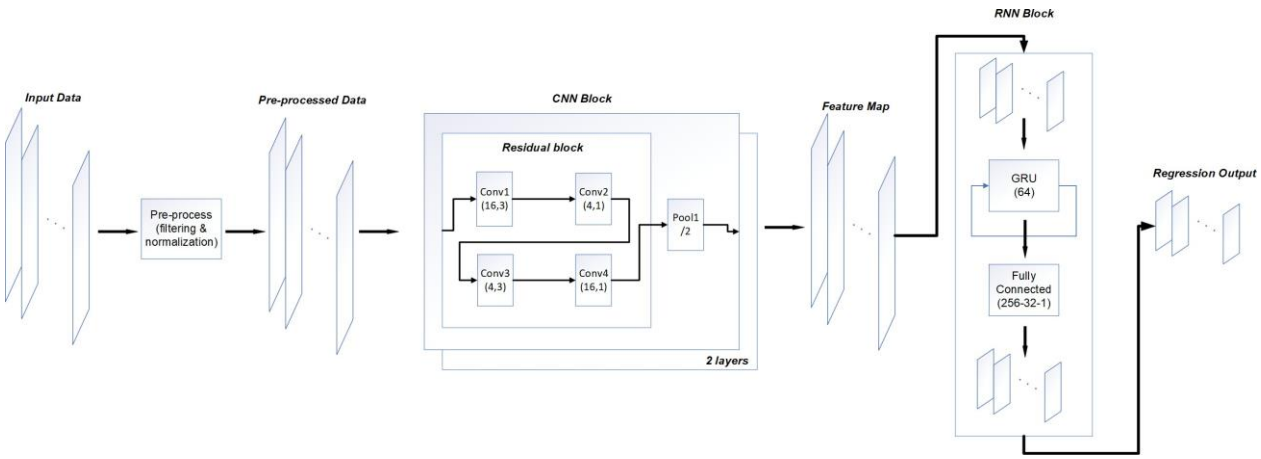

Fig. 5. The complete architecture of the proposed model.

TABLE I. THE RMSE FOR THE 10 SUBJECTS IN PH=30 MINUTES

Patient	Adult1	Adult2	Adult3	Adult4	Adult5	Adult6	Adult7	Adult8	Adult9	Adult10
Best	8.15	7.85	9.17	8.27	7.68	8.40	9.77	8.04	7.66	8.06
Best Avg	8.3050 ±0.64									

IV. Results

The root mean squared error (RMSE) between the predicted and given glucose results in PH=30 minutes is used to evaluate the performance of our model. The calculation of this assessment indicator can be expressed as

$$RMSE = \sqrt{\frac{1}{N}\sum_{m=1}^{N}(x(m) - \hat{x}(m|m-PH))^2} \quad (5)$$

where N denotes the data size, x refers to the given value and $\hat{x}(m|m-PH)$ refers to the prediction by the model.

The RMSE for each subject has been recorded and show in Table I. The proposed model achieves the best performance on subject Adult9 with the lowest RMSE (7.66). The average of the best results for each subject is 8.3050 and the standard deviation is 0.64.

TABLE II: Comparisons of the Prediction Performance between Different Models

Model	RMSE (best avg)
The proposed CRNN	8.3050±0.64
SVR [2]	20.3590±3.59
NNPG [3]	15.6020±1.94
Bidirectional RNN [12]	9.2360±0.86

To make the prediction performance of our model more convincible, a comparison table (refers to Table II) is created to compare the different prediction results between various algorithms.

V. Discussion

From Table I, it can be seen that except for the Adult3 and Adult7 subjects, the best RMSEs for other subjects are about 8 and even smaller than 8. This suggests that the proposed predictive algorithm experiences decent fitting.

From the comparison table (Table II), our model shows competitive performance among various models, such as SVR, NNPG and bidirectional RNN. It is worth mentioning that for comparability, the RMSE in the table is obtained by averaging the best results of 10 patients, and the parameters such as the dropout probability in the hidden layers are set to be the same for both bidirectional RNN and the suggested model. Since the mentioned parameters are not included in SVR and NNPG architecture, the results can be compared directly after running the codes. To be more specific, the RMSE result of our algorithm is about 60% better than that of SVR and around 47% better than that of NNPG. Even though the performance of bidirectional RNN is also prominent, our model is around 17ms/step faster than it in terms of training speed (our model: ≈48ms/step; bidirectional RNN: ≈65ms/step).

To visualize the comparison results and show the fitting degree, the predictive curve and the reference data in PH=30 minutes are plotted in Fig. 6. By looking at the plot, it is proved that the proposed CRNN model (the red curve) fits the true data (the orange curve) well.

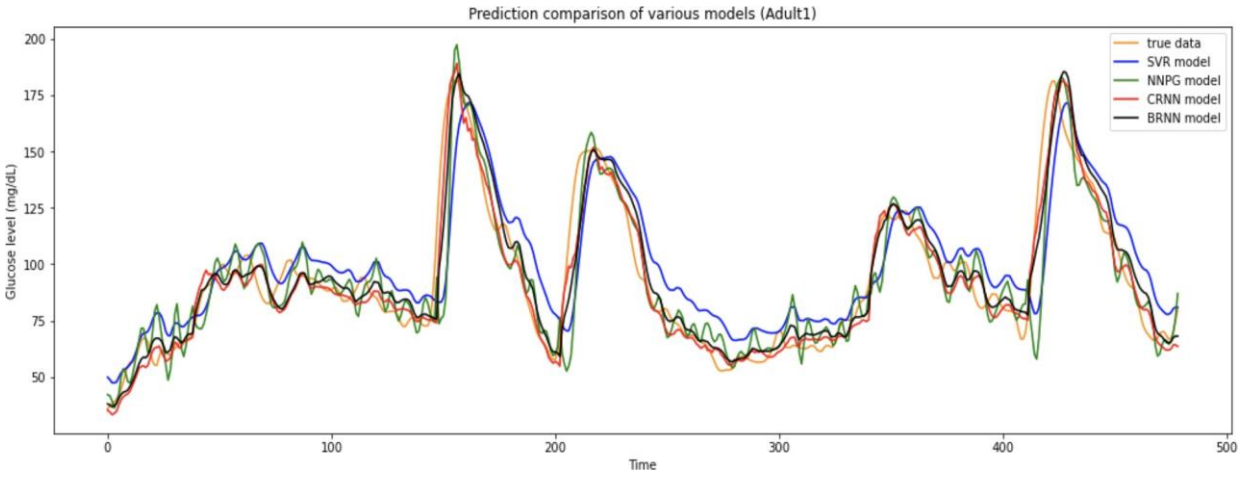

Fig. 6. Prediction comparisons of various models.

The superior performance of our model confirms the feasibility and correctness of the method adopted for glucose prediction. To confirm the applicability of each component of the method separately, the effects of using a component and not using a component are compared on a certain dataset. Here, we use Adult1 as an example. Under the condition that parameters remain unchanged, the RMSE changes from 8.31 to 18.56 when residual blocks are removed. This change proves that the residual block can indeed solve the degradation problem of deeper networks and improve the accuracy of the model. When deleting the dilation in the CNN part, the RMSE is 8.40, which is approximately 0.25 higher than that of the original one. This also demonstrates that a dilated CNN with a larger receptive field is important for great performance. And the RMSE of the model increases to 8.62 after changing GRU to LSTM, which implies the advantage of GRU. These RMSE differences under various circumstances are also shown in a comparison table (refers to Table III).

However, the proposed model can be improved in several aspects. For example, the architecture of our model is relatively simple with 20 convolutional layers, while the effect of residual can be more significant when the model gets much deeper. And since the data for training the model are generated from a simulator, more useful and valuable data such as real clinical data can be included to make the proposed predictive model more convincing.

TABLE III: Comparisons of RMSE Results under Different Circumstances (for Adult1)

CRNN with residual blocks	CRNN without residual blocks
8.15	18.56
CRNN with dilations	CRNN without dilations
8.15	8.40
CRNN with GRU cells	CRNN with LSTM cells
8.15	8.62

VI. Conclusion

In this paper, a novel deep learning model for glucose prediction is proposed. It is developed by combining both CNN and RNN: the multi-layer dilated causal CNN to extract patterns from the time series input, and the followed RNN including 64 GRU cells and 3 fully connected layers to predict the glucose level after 30 minutes by processing both the previous and current inputs. 10 virtual subjects with type 1 diabetes are employed to provide 10 simulated datasets with 6 fields of data, and the model is trained on these 10 datasets individually.

The average of the best RMSE of the 10 subjects is 8.3050 with a standard deviation of 0.64. This result shows the competitiveness of the suggested predictive algorithm among several common models. However, the work can be improved by involving much deeper layers and considering more clinical datasets in the future.

Appendix

Details of the 20-layer dilated causal CNN with residual blocks is shown in Fig. 7.

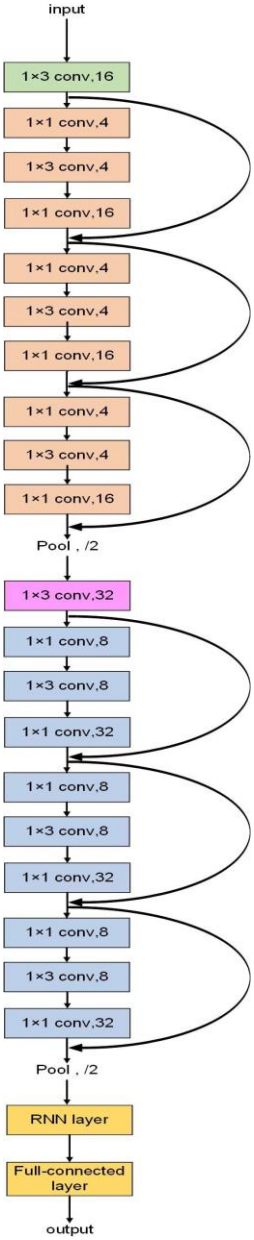

Fig. 7. The multi-layer dilated causal CNN with residual blocks.

Conflict of Interest

The author declares no conflict of interest.

Author Contributions

The author analyzed the dataset, built and trained the proposed model, evaluated the model performance, and wrote the paper.

References

[1] M. A. Atkinson, G. S. Eisenbarth, A. W. Michels, "Type 1 diabetes," *The Lancet*, vol. 383, issue 9911, pp. 69-82, ISSN 0140-6736, 2014.

[2] 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.

[3] C. Pe érez-Gand á, A. Facchinetti, G. Sparacino, C. Cobelli, E. J. Go ómez, M. Rigla, A. de Leiva, and M. E. Hernando, "Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring," *Diabetes Technology and Therapeutics*, vol. 12, no. 1, pp. 81-88, Jan, 2010.

[4] C. Zecchin, A. Facchinetti, G. Sparacino, G. D. Nicolao, and C. Cobelli, "Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1550–1560, Jun. 2012.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature 521*, pp. 436–444, May. 2015.

[6] H. N. Mharskar, S. V. Pereverzyev, and M. D. V. D. Walt, "A deep learning approach to diabetic blood glucose prediction," *Frontiers in Applied Mathematics and Statistics*, vol. 3, pp. 14, ISSN. 2297-4687, 2017.

[7] T. Zhu, K. Li, P. Herrero, J. Chen, and P. Georgiou, "A deep learning algorithm for personalized blood glucose prediction," in *Proc. KHD@ 27th International Joint Conference on Artificial Intelligence* (IJCAI), July 2018, pp. 74-78.

[8] J. Xie and Q. Wang, "Benchmarking machine learning algorithms on blood glucose prediction for type I diabetes in comparison with classical time-series models," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 11, pp. 3101-3124, Nov. 2020.

[9] K. Li, J. Daniels, C. Liu, P. Herrero, and P. Georgiou, "Convolutional Recurrent Neural Networks for Glucose Prediction," in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 603-613, Feb. 2020.

[10] A. V. D. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior and K. Kavukcuoglu, "WaveNet: A Generative Model for Raw Audio," *arXiv preprint, arXiv:1609.03499*, 2016.

[11] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," *arXiv preprint, arXiv:1512.03385*, 2015.

[12] Q. Sun, M. V. Jankovic, L. Bally, S. G. Mougiakakou, "Predicting Blood Glucose with an LSTM and Bi-LSTM Based Deep Neural Network," in *14th Symposium on Neural Networks and Applications (NEUREL)*.IEEE,pp. 1-5, 2018.

[13] M. E. Pennant, L. J. C. Bluck, M. L. Marcovecchio, B. Salgin, R. Hovorka, and D. B. Dunger, "Insulin administration and rate of glucose appearance in people with type 1 diabetes," *Diabetes Care*, vol. 31, no. 11, pp. 2183-2187, Nov. 2008.

[14] H. Cheng, Z. Xie, Y. Shi, and N. Xiong, "Multi-step data prediction in wireless sensor networks based on one-dimensional CNN and bidirectional LSTM," in *IEEE Access*, vol. 7, pp. 117883-117896, 2019.

[15] J. Chung, C. Gulcehre, K. Cho, and Y. Benjio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," *arXiv: Neural and Evolutionary Computing, arXiv:1512.03385*, 2014.

[16] W. Wang, N. Yang, F. Wei, B. Chang, and Z. Ming, "Gated self-matching networks for reading comprehension and question answering," in *Proc. the 55th Annual Meeting of the Association for Computational Linguistics*, 2017, vol. 1: Long Papers, pp. 189-198.

[17] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[18] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv: Machine Learning, arXiv:1609.04747*, 2016.

Yuxi Zhang was born on August 26, 2000, in Chongqing, China. She studied in computer science and technology at Xi'an Jiaotong-liverpool University in Suzhou, China during 2018-2020. Currently she is pursuing her degree of computer science and electrical engineering BEng (Hons) at the University of Liverpool, UK.