ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-6654-0540-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICASSP43922.2022.9746129

JOINT HYPOGLYCEMIA PREDICTION AND GLUCOSE FORECASTING VIA DEEP MULTI-TASK LEARNING

Mu Yang[1*], *Darpit Dave*[2], *Madhav Erraguntla*[2], *Gerard L. Cote*[2] *and Ricardo Gutierrez-Osuna*[2]

[1]University of Texas at Dallas, Richardson, Texas, USA
[2]Texas A&M University, College Station, Texas, USA
[1]mu.yang@utdallas.edu, [2]{darpitdave, merraguntla, gcote, rgutier}@tamu.edu

ABSTRACT

We present a multitask learning approach to the problem of hypoglycemia (HG) prediction in diabetes. The approach is based on a state-of-the-art time series forecasting model, N-BEATS, and extends it by adding a classification task so that the model performs both glucose forecasting (i.e., predicting future glucose values) and HG prediction (i.e., probability of future HG events sometime within the prediction horizon). We also propose an alternative loss function that penalizes forecasting errors in the HG range. We evaluate the approach on a dataset containing over 1.6M recordings from 112 patients with type 1 diabetes who wore a continuous glucose monitor (CGM) for 90 days. Our results show that the classification branch significantly outperforms the forecasting branch on the problem of HG prediction, and that the new loss function is more effective at reducing forecasting errors in the HG range than multi-task learning.

***Index Terms*—** hypoglycemia prediction, multitask learning.

1. INTRODUCTION

Patients with diabetes must manage two risks, those arising from sustained high blood glucose levels (hyperglycemia), which can lead to severe long-term complications (heart disease, kidney, eye and nerve damage), and those associated with glucose levels falling too low (hypoglycemia), which in the short-term can lead to seizures, loss of consciousness and even death [1]. Hypoglycemia (HG) is particularly problematic in type 1 diabetes, since patients must use insulin to prevent high glucose levels, which can lead to overdosing. Hence, accurately tracking and predicting glucose is a critical component in diabetes management. This can be facilitated by using a continuous glucose monitor (CGM), a wearable device that can measure glucose continuously every 5-15 minutes for up to 2 weeks [2, 3]. In addition to using CGMs, algorithms can also be used to predict future glucose values from the history of recent CGM measurements [4, 5]. This is known as HG prediction, and is the subject of this paper.

Current work on HG prediction can be grouped into two classes: time-series-forecasting (TSF) and classifier-based. TSF methods predict a single future glucose value at a specific time point [6, 7] or multiple glucose values within a prediction horizon [8, 9, 10]. If the predicted glucose value is below a threshold, the system can alert patients of an impending HG event so they can act (e.g., drink fruit juice). TSF models are generally optimized on the overall blood glucose range. However, HG events are uncommon (1-2 per week, on average), so TSF can be error-prone in the HG range (<70 mg/dL), where predictions matter most. The second type of technique, classification methods, directly predict the probability that a HG event will occur in the near future, without attempting to predict the exact glucose value [11, 12]. From a computational standpoint, classification problems are easier and more efficient to solve than TSF, and can be more accurate for HG prediction [11]. However, a classifier provides relatively limited information compared to TSF.

To address this issue, we propose a multitask learning framework that unifies these two tasks (TSF and classification). Our framework is based on the N-BEATS architecture, a state-of-the art deep-learning model for TSF that has been recently used for glucose forecasting [10] to enter (and win) the 2020 Blood Glucose Level Prediction Challenge [13]. We add an auxiliary branch to N-BEATS that performs classification, so our model learns to simultaneously forecast future glucose values and predict the probability of a future HG event. Multitask learning allows information to be shared within the two tasks, and can potentially benefit both individually. Specifically, we hypothesize that supervision from the classification task can help the TSF branch make more accurate glucose forecasting in the HG range. Further, to encourage the TSF branch to produce more accurate glucose predictions in the HG range, we replace the mean square error (MSE) loss function of N-BEATS, with a Normalized MSE (NMSE). This modification can significantly improve the TSF branch in the HG range. We evaluate the system against the original N-BEATS system and a strong feature-based classi-

The first author performed this work while at Texas A&M University.

This work was supported by PATHS-UP (NSF ERC), NSF award 2037383, and a grant from the Kleberg Foundation.

fication baseline [12]. We argue that this multitask learning design can be more informative in clinical settings by integrating predictions from both streams.

2. BACKGROUND AND RELATED WORK

Extensive prior work has been conducted on predicting short-term outcomes from CGMs, with TSF-based methods having drawn more attention than classifier-based methods. Early work in glucose forecasting include Autoregressive Integrated Moving Average (ARIMA) [14], and statistical machine learning methods such as support vector regression [15]. More recently, in light of the success of deep learning, methods based on Deep Neural Networks (DNN) [16, 17, 18] have achieved promising performance. To a lesser extent, classifier-based approaches have also been studied for HG prediction. Sudharsan et al. [11] proposed a simple yet effective Random Forest (RF) classifier for HG prediction. Also recently, our group [12] proposed a set of hand-crafted features to enhance the RF classifier. Galadeta et al. [6] compared various classification-based and TSF-based algorithms. However, their comparison did not include DNNs.

3. PROBLEM STATEMENT

Given N previous glucose measurements $\mathbf{x_{t-N-1:t}} = [x_{t-N-1}, \ldots, x_t] \in \mathbb{R}^N$ until time t , the goal of HG prediction is to estimate whether a HG event will occur in the near future. Prior studies seek to predict HG at a particular future time point [6, 7] (e.g., in exactly 30 minutes), while others seek to predict HG within a prediction horizon [8, 10, 12, 19] (e.g., sometime in the next 30 minutes). In this work, we consider the latter case, as in clinical applications the risk at an exact future time is less important than the overall risk status in the entire prediction horizon [12]. That is, our goal is to predict the HG label $y_{t+1:t+PH}$, where PH denotes the prediction horizon. Specifically, $y_{t+1:t+PH} = 1$ if any of the glucose values in $[x_{t+1}, \ldots, x_{t+PH}]$ is below a threshold, and $y_{t+1:t+PH} = 0$ otherwise [12].

TSF-based methods aim at predicting $\mathbf{x_{t+1:t+PH}}$, i.e. the glucose values in the prediction horizon, from $\mathbf{x_{t-N-1:t}}$. From these, the predicted HG label $\hat{y}_{t+1:t+PH}$ can be obtained from the predicted glucose values $\mathbf{\hat{x}_{t+1:t+PH}}$:

$$\hat{y}_{t+1:t+PH} = \begin{cases} 1, & \text{if } \min(\mathbf{\hat{x}_{t+1:t+PH}}) < c_g \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where c_g is a glucose level threshold (e.g., 70 mg/dL).

In contrast, classification methods cast HG prediction as a binary classification problem and directly output the probability distribution $p(\hat{y}_{t+1:t+PH})$, from which the predicted HG label $\hat{y}_{t+1:t+PH}$ is decided as follows:

$$\hat{y}_{t+1:t+PH} = \begin{cases} 1, & \text{if } p(\hat{y}_{t+1:t+PH} = 1) > c_p \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

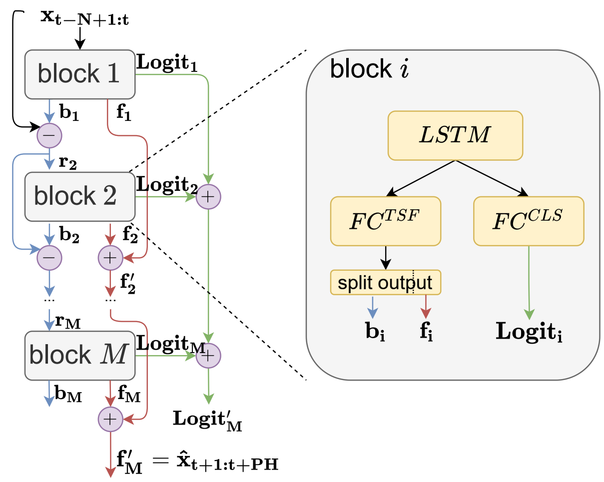

Fig. 1. Architecture of the proposed multitask learning system. Block structure of N-BEATS (left). A classification branch is added to each block (right).

where c_p is a probability threshold for the positive class (i.e., HG). Following [12], we consider a 4-hour CGM history, and a prediction horizon of 30 mins. With a CGM recording interval of 5 mins in our dataset (see section 5), this results in $N = 48$ and $PH = 6$.

4. METHODS

Our proposed multitask learning model is based on N-BEATS [10]. N-BEATS consists of a stack of M DNN blocks. Each block i outputs a backcast $\mathbf{b_i} \in \mathbb{R}^N$ which tries to reconstruct the input signal, and a forecast $\mathbf{f_i} \in \mathbb{R}^{PH}$ which tries to predict and match the target signal. In each block, the final hidden state of the LSTM is projected by a Fully Connected (FC) layer FC^{TSF}, and then split into $\mathbf{b_i}$ and $\mathbf{f_i}$, as shown in Fig. 1. The backcast $\mathbf{b_i}$ is subtracted from the block input to form a residual $\mathbf{r_{i+1}} = \mathbf{r_i} - \mathbf{b_i}$, which functions as the input signal to the next block (Note that $\mathbf{r_1} = \mathbf{x_{t-N-1:t}}$). This encourages the model to learn to reconstruct part of the input signal and predict part of the target signal block by block. The forecast $\mathbf{f_i}$ is summed to construct the final forecasting $\mathbf{\hat{x}_{t+1:t+PH}} = \mathbf{f'_M} = \sum_{i=1}^{M} \mathbf{f_i}$. To train the model, the most important loss functions are reconstruction loss and forecasting loss. The block-wise reconstruction loss L_i^r and block-wise forecasting loss L_i^f are defined as:

$$L_i^r = \text{MSE}(\mathbf{b_i}, \mathbf{r_i}) \tag{3}$$

$$L_i^f = \text{MSE}(\mathbf{f'_i}, \mathbf{x_{t+1:t+PH}}) \tag{4}$$

where MSE denotes Mean Square Error, and $\mathbf{f'_i} = \sum_{k=1}^{i} \mathbf{f_k}$ represents the sum of the partial forecasts until block i. The final reconstruction and forecasting losses are weighted by block depth and added, respectively. The total loss L^{TSF} is computed by the reconstruction loss, forecasting loss and an auxiliary loss. Interested readers can refer to [10] for more details.

4.1. Extension to multitask learning

As shown in Fig. 1, to allow N-BEATS to perform classification, we use a FC layer FC^{CLS} to project the last hidden state from the LSTM to the binary $Logit_i$ for the two classes (HG and non-HG). This classification branch and the TSF branch share the same LSTM encoder, which encourages the learning of common information in the two tasks. From the perspective of the entire multitask learning framework, the final logit $Logit'_M = \sum_{i=1}^{M} Logit_i$ forms the output from the classification branch, while $\mathbf{f'_M} = \sum_{i=1}^{M} \mathbf{f_i}$ forms the output from the TSF branch. The classification loss L^{CLS} is computed by *Cross Entropy* (CE):

$$L^{CLS} = \text{CE}(Logit'_M, y_{t+1:t+PH}) \tag{5}$$

The total loss of the multitask learning framework is a weighted sum of L^{TSF} and L^{CLS}:

$$L = L^{TSF} + \lambda L^{CLS} \tag{6}$$

where λ is a scalar.

4.2. Normalized Mean Square Error

The original TSF task in N-BEATS penalizes errors uniformly over the entire glucose range. However, HG events are rare, so errors in that range contribute less to the overall cost. This can lead to sub-optimal predictions in the HG range, where errors matter most. To address this issue, we replace the conventional MSE with a Normalized MSE (NMSE):

$$\text{NMSE} = \frac{1}{L} \sum_{i=1}^{L} \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2 \tag{7}$$

where $y_1, \ldots, y_L \in \mathbb{R}$ represent the reference values, and $\hat{y}_1, \ldots, \hat{y}_L$ are the corresponding predicted values. Adding the reference value y_i to the denominator places a larger penalty on the HG range, where glucose values are small.

5. EXPERIMENTAL SETUP

We evaluated the multi-task architecture [1] on the dataset reported in [9], which includes over 1.6 million glucose recordings from 112 patients. Of those readings, only 2.13% were in the HG range (<70 mg/dL). Following [12], we train subject-independent models. That is, we partition CGM readings for each patient as 80/10/10 for training/validation/test respectively, following temporal order. The final training/validation/test sets are constructed by merging the corresponding partition from all patients. We also use time of day as an input, computed by sine encoding the time stamps of the CGM readings. We find that adding the time encoding slightly improves performance (results not included).

[1]Code is available at `https://github.com/Mu-Y/HG_prediction`

Method	Full range	Hypoglycemia range
NB-tsf	24.09	38.02
MT-NB-tsf	**23.73**	37.43
MT-NB-L*-tsf	25.70	**29.21**

Table 1. RMSE for the TSF methods on test set.

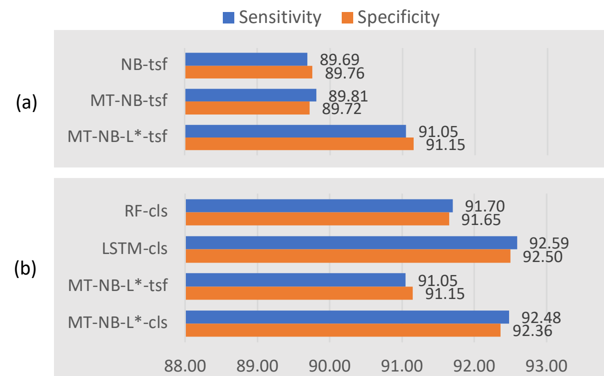

Fig. 2. (a) Effect of multitask learning and NMSE loss. (b) Comparison between baselines and the proposed multitask learning model.

For convenience, we use the following notation for the models in our study: (1) ***NB-tsf***: the N-BEATS model in [10], where *tsf* indicates that the model only performs the TSF task; (2) ***MT-NB***: our proposed multitask model in Section 4.1; (3) ***MT-NB-L****: the multitask model trained with the NMSE loss in Section 4.2, with the suffix *-tsf* and *-cls* representing which branch is used to decode the HG prediction. We use a FC^{CLS} with 300-unit hidden linear layer and 2-unit output layer, with ReLU activation. We set $\lambda = 100$. These hyperparameters were tuned based on validation set performance. Other hyperparameters follow [10]. We use the validation set for early stopping, with patience=10 epochs.

We perform two evaluations: (1) glucose forecasting and (2) HG prediction. Following prior work [6, 8, 12], we use Root Mean Square Error (RMSE) to evaluate the former, and Sensitivity and Specificity for the latter. To deliver HG predictions from the TSF branch, we adjust the threshold c_g in eq. (1) such that Sensitivity and Specificity are matched. For the classifier branch, the HG probability is computed by Softmax: $p(\hat{y}_{t+1:t+PH}) = \text{Softmax}(Logit'_M)$. Then, we adjust the threshold c_p in eq. (2) in the same way as c_g.

6. RESULTS

6.1. Improvements on the TSF task

First, we evaluate the three methods on the TSF task, in terms of RMSE on the test set. Following [10], we use the glucose value at the final time step of the prediction horizon to compute RMSE. Results are summarized in Table 1, in terms of the overall RMSE and the RMSE in the HG range. For the baseline model (*NB-tsf*), we observe a large difference in RMSE between the full range (24.09) and the HG range (38.02), which indicates that the model is performing sub-optimally in the HG range, where prediction errors are critical. Comparing *MT-NB-tsf* with *NB-tsf*, we also observe a modest improvement in glucose forecasting for the multi-tasking model, which shows that the classification task can benefit the TSF task. However, the largest improvements in prediction in the HG range occur as a result of incorporating the NMSE loss into *MT-NB-tsf*, for a reduction from 37.43 to 29.21. This significant RMSE reduction suggests that the NMSE loss can effectively improve the accuracy of the TSF task in the HG range. On the flip side, using the NMSE loss leads to an increase in RMSE for the full range, though the increase (1.97 mg/dL) is likely not meaningful clinically when glucose is normal or elevated.

Fig. 2(a) shows the results of HG prediction (i.e., Sensitivity and Specificity). *MT-NB-L*-tsf* significantly outperforms the other two models, which corroborates the results on the glucose TSF task. Since models *MT-NB-tsf* and *NB-tsf* perform comparably (37.43 vs. 38.02), it appears that the improvements in HG prediction for the *MT-NB-L*-tsf* model are due to the NMSE loss rather than from multitasking.

6.2. TSF vs classification for hypoglycemia prediction

Next, we asked whether the TSF branch or the classification branch should be preferred for HG prediction, by comparing the Sensitivity and Specificity of *MT-NB-L*-cls* and *MT-NB-L*-tsf*. We also included two baselines to the comparison: *RF-cls*, a feature-based Random Forest (RF) classifier with the optimal set of features reported in [12], and *LSTM-cls*, a model without the TSF branch in the multitask learning model, which serves as a single-task counterpart to *MT-NB-L*-cls*. Results are shown in Fig. 2(b). The three classification methods (*MT-NB-L*-cls*, *LSTM-cls*, and *RF-cls*) outperform the TSF method (*MT-NB-L*-tsf*), which indicates that a classifier is superior to a TSF model in terms of HG prediction. Note also that *MT-NB-L*-cls* performs comparably to its single-task classification counterpart *LSTM-cls*, which suggests that adding a TSF task to a classification task does not affect the performance of the latter. Finally, we find that the DNN-based classification methods (*LSTM-cls* and *MT-NB-L*-cls*) also outperform the feature-based RF, thanks to the strong representation-learning ability of DNNs.

7. DISCUSSION AND FUTURE WORK

We have presented a DNN-based multitask learning framework that unifies glucose forecasting and hypoglycemia prediction. By adding a classification task, the TSF branch can outperform its single-task counterpart, but a more significant improvement in glucose forecasting is achieved by using the NMSE loss. Further, classifier methods outperform TSF methods in terms of HG prediction. One reason is that the supervision of the classification task is directly related to the final HG prediction evaluation, making the classification task easier and advantageous compared to the TSF task.

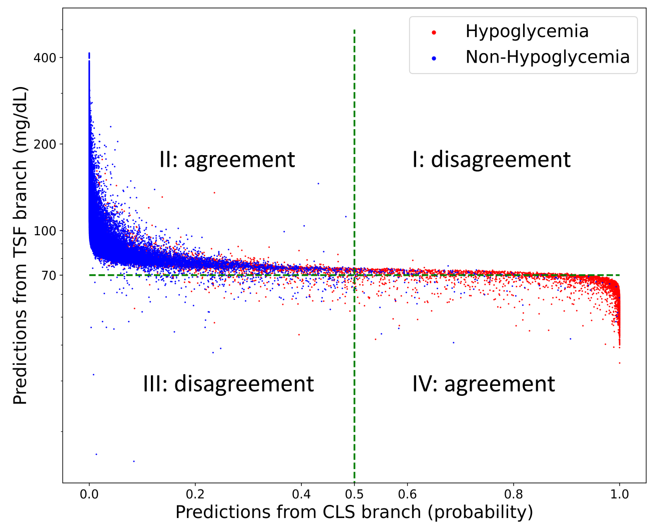

Fig. 3. Predictions from the two branches of the multitask learning model. For the TSF branch, we take the minimum of the predicted glucose values within PH. For the CLS branch, we simply show the probability of HG. Red points represent true instances of HG.

Fig. 3 shows a scatter plot with outputs from both branches. Predictions in Quadrants II and IV represent the majority of cases, where the TSF branch and classification branch outputs agree (both negative or both positive). Quadrants I and III are cases where two branches make conflicting predictions. Although the classification branch produces more accurate HG prediction, relying merely on the probability output may still lead to ignoring HG events (e.g. red points in Quadrant III). Taking TSF predictions into consideration can aid in such cases and improve interpretability by providing the patient's overall glucose trajectory in the prediction horizon, which can be used to inform more cautious treatment. A unified multitask learning model offers the flexibility of choosing a specific branch for training, and thus may simplify the pipeline design.

In future work, we will examine methods to combine predictions from the two branches. For example, these may be combined using logic gates (AND, OR), a weighted average relative to the performance of each branch, or in a stacked generalization fashion by using a separate classifier (a FC layer) that uses both branch outputs as input and predicts the probability of HG. Other ensemble approaches such as max-margin classifier or jointly training the ensembler with the multitask learning model could also be explored.

8. REFERENCES

[1] Mazen Alsahli and John E Gerich, "Hypoglycemia.," *Endocrinology and metabolism clinics of North America*, vol. 42, no. 4, pp. 657–676, 2013.

[2] David Rodbard, "Continuous glucose monitoring: a review of recent studies demonstrating improved glycemic outcomes," *Diabetes technology & therapeutics*, vol. 19, no. S3, pp. S–25, 2017.

[3] Giacomo Cappon, Martina Vettoretti, Giovanni Sparacino, and Andrea Facchinetti, "Continuous glucose monitoring sensors for diabetes management: a review of technologies and applications," *Diabetes & metabolism journal*, vol. 43, no. 4, pp. 383–397, 2019.

[4] Omer Mujahid, Ivan Contreras, and Josep Vehi, "Machine learning techniques for hypoglycemia prediction: Trends and challenges," *Sensors*, vol. 21, no. 2, pp. 546, 2021.

[5] Virginie Felizardo, Nuno M Garcia, Nuno Pombo, and Imen Megdiche, "Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction–a systematic literature review," *Artificial Intelligence in Medicine*, p. 102120, 2021.

[6] Matteo Gadaleta, Andrea Facchinetti, Enrico Grisan, and Michele Rossi, "Prediction of adverse glycemic events from continuous glucose monitoring signal," *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 650–659, 2018.

[7] Md Fazle Rabby, Yazhou Tu, Md Imran Hossen, Insup Lee, Anthony S Maida, and Xiali Hei, "Stacked lstm based deep recurrent neural network with kalman smoothing for blood glucose prediction," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, pp. 1–15, 2021.

[8] Clara Mosquera-Lopez, Robert Dodier, Nichole Tyler, Navid Resalat, and Peter Jacobs, "Leveraging a big dataset to develop a recurrent neural network to predict adverse glycemic events in type 1 diabetes," *IEEE journal of biomedical and health informatics*, 2019.

[9] Ian Fox, Lynn Ang, Mamta Jaiswal, Rodica Pop-Busui, and Jenna Wiens, "Deep multi-output forecasting: Learning to accurately predict blood glucose trajectories," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 1387–1395.

[10] Harry Rubin-Falcone, Ian Fox, and Jenna Wiens, "Deep residual time-series forecasting: Application to blood glucose prediction," in *Proceedings of the 5th International Workshop on Knowledge Discovery in Healthcare Data co-located with 24th European Conference on Artificial Intelligence, KDH@ ECAI*, 2020, pp. 29–30.

[11] Bharath Sudharsan, Malinda Peeples, and Mansur Shomali, "Hypoglycemia prediction using machine learning models for patients with type 2 diabetes," *Journal of diabetes science and technology*, vol. 9, no. 1, pp. 86–90, 2014.

[12] Darpit Dave, Daniel J DeSalvo, Balakrishna Haridas, Siripoom McKay, Akhil Shenoy, Chester J Koh, Mark Lawley, and Madhav Erraguntla, "Feature-based machine learning model for real-time hypoglycemia prediction," *Journal of Diabetes Science and Technology*, p. 1932296820922622, 2020.

[13] Cindy Marling and Razvan Bunescu, "The ohiot1dm dataset for blood glucose level prediction: Update 2020," in *CEUR workshop proceedings*. NIH Public Access, 2020, vol. 2675, p. 71.

[14] Meriyan Eren-Oruklu, Ali Cinar, and Lauretta Quinn, "Alarms for continuous glucose monitors: Hypoglycemia prediction with subject-specific recursive time-series models," *Journal of Diabetes Science and Technology*, vol. 4, no. 1, pp. 25, 2010.

[15] Eleni I Georga, Vasilios C Protopappas, Diego Ardigo, Demosthenes Polyzos, and Dimitrios I Fotiadis, "A glucose model based on support vector regression for the prediction of hypoglycemic events under free-living conditions," *Diabetes technology & therapeutics*, vol. 15, no. 8, pp. 634–643, 2013.

[16] Kezhi Li, Chengyuan Liu, Taiyu Zhu, Pau Herrero, and Pantelis Georgiou, "Glunet: a deep learning framework for accurate glucose forecasting," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 414–423, 2019.

[17] John Martinsson, Alexander Schliep, Björn Eliasson, and Olof Mogren, "Blood glucose prediction with variance estimation using recurrent neural networks," *Journal of Healthcare Informatics Research*, vol. 4, no. 1, pp. 1–18, 2020.

[18] Hadia Hameed and Samantha Kleinberg, "Comparing machine learning techniques for blood glucose forecasting using free-living and patient generated data," in *Machine Learning for Healthcare Conference*. PMLR, 2020, pp. 871–894.

[19] Touria El Idrissi and Ali Idri, "Deep learning for blood glucose prediction: Cnn vs lstm," in *International Conference on Computational Science and Its Applications*. Springer, 2020, pp. 379–393.