A Performance Analysis of RNN Algorithms for Accurate Hypoglycemia Prediction in Type-1 Diabetes

Sunandha Rajagopal[1], Dr. N. Thangarasu[2]

[1]*Research Scholar, Karpagam Academy of Higher Education, Coimbatore, sunandhar@gmail.com*
[2]*Assistant Professor, Karpagam Academy of Higher Education, Coimbatore, drthangarasu.n@kahedu.edu.in*

KEYWORDS

RNN, Algorithms for Accurate

ABSTRACT

Diabetes is a chronic disease originating from preeminent blood glucose levels where there is an impairment in the ability of the human body to produce or effectively use insulin, therefore resulting in possible complications to various organ systems. Hypoglycemia can be defined as a critical condition in diabetes that is characterized by very low levels of blood sugar and can emanate from any imbalance between insulin, glucose, and external factors such as medication or physical activity. The prediction of hypoglycemia is a very important task in diabetes management, which encompasses sophisticated technologies and deep-learning systems, such as RNNs, in performing analyses on patient-specific data. These must provide timely warnings to avoid hazardous blood sugar dips. This paper takes a close look at the performance of RNN in the case of accurate hypoglycemia prediction among patients with Type-1 Diabetes, considering a dataset from Shanghai T1DM.

In this work, three different RNN architectures are considered for performance evaluation: Long-Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Simple RNN. The main goal was to compare their predictive performances in forecasting hypoglycemic events, which is an issue of utmost relevance when it comes to proactive diabetes management. The results show generally variable performances by the RNN models. Overall, GRU performed with striking accuracy in hypoglycemia predictions, while LSTM had high specificity. These findings underline that various metrics should be considered for the comprehensive evaluation of predictive models in the management of diabetes. It will give an idea about various RNN algorithms, strengths, and weaknesses to develop more effective and personalized strategies related to hypoglycemia prediction in Type-1 Diabetes.

1. Introduction

In Type-1 Diabetes is and autoimmune disorder in which, the immune system, attacks the beta cells and kills them by mistake. Insulin, an important hormone in blood sugar regulation, becomes insufficient. Consequently, glucose cannot get inside the body's cells, raising blood sugar levels. But while certain specific factors that provoke this autoimmune response are unclear, genetic and environmental factors are known to bear some causative relationship to this disorder. Unlike type 2 diabetes, which usually arises later in life, type 1 diabetes normally arises very early in childhood or even adolescence. Diabetes, if not treated, especially type 1, may lead to serious and life-threatening circumstances.

Untreated or poorly managed diabetes may result in Diabetic Keto Acidosis, which is very dangerous because of the accumulation of ketones in the blood, which may result in dehydration, vomiting, and may result in diabetic coma. On the other hand, hypoglycemia is the condition that results in an extremely low amount of blood sugar. For instance, this will happen if a diabetic person has very high insulin levels, skips meals, or is highly engaged in strenuous physical exercise without proper adjustments in the insulin dose. Severe hypoglycemia is dangerous and can go further to implicate itself through confusion, seizures, or unconsciousness. To avoid such complications, people with diabetes need to maintain their blood glucose by daily monitoring of blood glucose, medication management, and lifestyle changes. New innovations like Continuous Glucose Monitoring (CGM) systems and insulin pumps have reformed diabetes care by contributing real-time insights of blood glucose values and allowing for more accurate dosing of insulin.

By being able to foresee hypoglycemia with the use of CGM data, one can implement timely interventions of adjusting the dosage of insulin or taking glucose-rich food to avoid severe complications. All these enable a person with diabetes to understand and manage their condition far better, which considerably reduces the risks of long-term complications, improving the quality of life. Therefore, if they can predict the onset of hypoglycemia by using newer digital diabetic management devices data, this will be crucial to diabetes management. This predictive capability can enable users to take corrective action concerning insulin intake or other precautionary measures well in time, greatly reducing the risk of serious hypoglycemia. The present paper carries out a detailed performance analysis of RNN algorithms with a view to their efficiency regarding correct prediction of hypoglycemic events for Type-1 diabetic patients.

Recognizing that there is a crucial need for timely intervention strategies that can prevent hypoglycemic episodes, our study investigates in great depth several RNN model capabilities running different prediction horizons of 60, 75, and 90 minutes. This work, by investigating predictive accuracy and reliability, aspires to add value to the ongoing pursuit of refining predictive tools for enhanced diabetes management. For the purpose of this study, we use a subset of the Shanhai T1DM dataset.

Related Works

Approaches toward hypoglycemia prediction in Type 1 diabetes try to bring improvements over patient management by creating models that could successfully anticipate low events of blood glucose. These initiatives seek proactive interventions, such as adjusting insulin dosages, with the intent of mitigating hypoglycemia risks. Even if there is remarkable development in the area of diabetic health care, challenges still exist due to the complex behavioural pattern of glucose, which demand personalized approaches. Other challenges include the continuous real-time monitoring that will contribute to maintaining data quality and availability, and also to engage users and improve adherence to the predictions from the models. The behaviours of patients are important in understanding how to use these predictive tools effectively.

In 2020, Dae-Yeon Kim et al. proposed a deep learning model, especially RNNs, for personalized glucose prediction of inpatients affected by type 2 diabetes. As mentioned, this has been done in an attempt to put forward something useful that would assist medical staff in monitoring blood glucose and maintaining appropriate doses of insulin. [8]. The study only focuses on the performance of simple RNNs, gated recurrent units(GRUs), and long short-term memory (LSTM) networks with GRU being the best among them all. To achieve this, data from continuous glucose monitoring devices for a week are utilized by the proposed model, which scored an average root mean squared error (RMSE) of 21.5 and mean absolute percentage error (MAPE) of 11.1% during twenty hospitalized type 2 diabetic patients care. This research does valuable service in Type-2 diabetes mellitus by addressing the gap between studies about Type-1 patients' work and providing an equally good performance as before. [9].

Alfian et al. presented the machine-learning approach for predicting blood glucose levels in Type 1 diabetes patient using their innovative model in the same year, 2020. This artificial neural network (ANN) includes time-domain features that are used to predict future concentration of glucose at intervals of 15, 30, 45 and 60 minutes by probing into each of the preceding thirty minutes of blood sugar readings. It enhances characteristics from previous 30-minute glucose data with temporal domain properties within its input data that increases throughputs. Unlike many other data-driven models such as Support Vector Regression (SVR), K-Nearest Neighbour (KNN), Random Forest (RF), Adaptive Boosting (AdaBoost), Decision Tree (DT) and eXtreme Gradient Boosting(xgboost), this ANN-based prediction model performs better when tested on type-1 diabetics, consisting of twelve patients. Therefore, this proposed model achieved a high performance level with average root mean square error values of RMSE=2.82 mg/dL for PH=5min., RMSE=6.31 mg/dL for PH=10 min., RMSE=10.65 mg/dL for PH=15 min., and RMSE=15.33 mg/dL for PH=20 min. The integration of time-domain attributes enhances predictive capability, enabling early alerts for preventive measures before critical hypoglycemic or hyperglycaemic events occur [10].

Ma, Yu, Yang, and Zhao introduced a hypoglycemia early alarm method for type 1 diabetes patients in 2022, which is based on multi-dimensional sequential pattern mining. Since T1D hypoglycemia, particularly nighttime hypoglycemia, is a priority due to its criticality, the study is centred on developing an efficacious early alarm system. A multi-dimensional database is generated to organize blood glucose, meal, and insulin time series, and real-time retrieval of hypoglycemia sequence patterns is conducted utilizing the UniSeq algorithm. Evaluating the OhioT1DM dataset, results show a sensitivity of 75.76%, precision of 75%, F1 score of 75.38%, and an early alarm time of 25.17 minutes. The study highlights the potential of multi-dimensional sequential pattern mining for comprehensive diagnosis support in personalized treatment, offering valuable insights and early warnings for effective blood glucose management in T1D patients [11].

In 2022, Stella Tsichlaki, Lefteris Koumakis, and Manolis Tsiknakis undertook a systematic review regarding cutting-edge approaches for monitoring and preventing hypoglycemic events in type one diabetic patients—this review sought to discuss various detection methods with emphasis on technological advances. They applied PRISMA recommendations in their search through PubMed, Google Scholar, IEEE Xplore and ACM Digital Library. In all, fifteen papers were found that had designed predictive models for T1D related hypoglycemia with each paper opting for different approaches like statistical (10%), machine learning (52%) and deep learning techniques (38%). The most widely used algorithms include Kalman filtering, support vector machines, k-nearest neighbors and random forests. Consequently; it is evident that the predictive models performed quite well with an accuracy range of between 70% to 99% as an indication of these technologies potential applications in predicting hypoglycemia in individuals with T1D. [12].

In their 2022 work, Zhu et al. introduced a new way of promoting self-management in type 1 diabetes (T1D) using wearables and deep learning algorithms. They used the CGM-in-a-wristband sensor fusion to develop the ARISES platform with a DL-based algorithm. This system exhibited an excellent average root mean square error (RMSE) of 35.28 ± 5.77 mg/dL for a sixty-minute future glucose level prediction horizon. It is worth noting that for hypoglycemia detection, the correlation coefficients were Matthews's [sic] 0.56 ± 0.07, while hyperglycemia was at 0.70 ± 0.05 respectively by the algorithm. The inclusion of wristband data significantly reduced the RMSE by 2.25 mg/dL ($p < 0.01$). An application on a smartphone running ARISES can provide real-time decision support, which creates a promising direction toward mitigating serious complications and improving T1D self-care [13].

In 2022, Peng, Li, Wang, and Yan evaluated blood glucose prediction and raised warning of hypoglycemia using two deep learning techniques LSTM and GRU on continuous glucose curves from 100 Diabetes Mellitus patients over 72 hours. The model demonstrated superior predictive capabilities with mean RMSE values of 0.259, 0.272, 0.275, and 0.278 (mmol/l) for prediction horizons of 15min, 30min, 45min, and 60min. These values were significantly lesser than those obtained with the LSTM and GRU models ($p < 0.001$). The model based on LSTM and GRU, constantly exhibited sensitivity and false-negative rates over time, with a sustained false-positive rate. The study highlights the effectiveness of the model by outstanding precision in blood glucose prediction and hypoglycemia warning, emphasizing its potential for enhancing diabetes management [14].

Thomsen et al. addressed hypoglycemia in T2D patients undergoing insulin treatment in 2023 using a deep-transfer learning approach. They employed CGM data of T1D and T2D patients, as well as labelled one-hour samples into positive or negative classes based on a hypoglycemic criterion. They first pre-trained CNN using T1D data and then fine-tuned the network further by training with T2D data optimized for AUC. It showed outstanding results when externally validated on a separate dataset of T2D, with an AUC of 0.941; the PPV was 40.49% at 95% specificity and 69.16% sensitivity was obtained for this deep transfer learning model. Therefore, this model may prove to be very promising for the accurate prediction of hypoglycemia, thereby enhancing the quality of life and mortality in insulin-treated T2D patients [15].

Our paper is focused on the performance analysis of the RNN algorithm family for accurate hypoglycemia prediction in Type-1 Diabetes. In this regard, three different RNN architectures are evaluated: LSTM, GRU, and Simple RNN. Contrary to related works, the key objective here is to investigate the predictive performance of these models. Most importantly, the study clearly reveals some nuanced performances across RNN architectures: the high precision by GRU and high specificity by LSTM.

2. Methodology

This study undertakes a comprehensive investigation on the accuracy of hypoglycemia prediction in Type-1 diabetes, employing Recurrent Neural Network (RNN) algorithms. This focuses on three different RNN architectures – Simple RNN, Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM) networks. In addition to this, to correspond to variability, the Monte Carlo simulation technique is also employed. So, the variations in the initial model conditions are trained, and the mean predictions from the simulations are finally considered.

A. Dataset

For the purpose of our study, we utilized a subset of the ShanghaiT1DM dataset originated from Shanghai, China. The dataset contains CGM data from 12 individuals diagnosed with Type-1 Diabetes Mellitus (T1DM) collected through 3 to 14 days. The data comprises of CGM values, blood glucose levels, insulin doses, carbohydrate intake, as well as age and gender. [1].

In order to suit the method of our study, the dataset was rearranged. Each instance in the newly arranged dataset now includes details on insulin dosage, age, CGM measurements, and carbohydrate intake at the time of the recorded event, along with subsequent CGM readings at 30, 45, and 60 minutes. For each target column, namely PH60, PH75, and PH90, the occurrence of hypoglycemia is identified by assessing whether the blood glucose level falls below 70. Events of hypoglycemia are marked as 1, otherwise 0.

To ensure privacy of the patients, the original Patient IDs are replaced with numbers ranging from 1001 to 1012. Furthermore, to ensure the effectiveness of our model, the dataset was partitioned into an 80% (8000 rows) training set for model development and a 20% (2000) testing and validation set to assess the model's generalization performance. Table 1 provides a glimpse into a sample of the dataset.

Table1: Sample dataset

Age	CGM	Insulin	Carbs	PH60	PH75	PH90	Hypo60	Hypo75	Hypo90
66	113.4	6	0	156.6	162	163.8	0	0	0
68	156.6	9	44	131.4	129.6	129.6	0	0	0
68	199.8	8	46	178.2	172.8	165.6	0	0	0
68	244.8	8	22	216	210.6	201.6	0	0	0
37	142.2	7	83	81	68.4	61.2	0	1	1
67	88.2	5	26	64.8	82.8	115.2	1	0	0
58	104.4	7	50	70.2	73.8	68.4	0	0	1
57	205.2	9	60	174.6	169.2	163.8	0	0	0

PatientId	1001	1002	1003	1004	1005	1006	1007	1008

B. Data Preprocessing

The dataset, includes vital features like Age, CGM readings, dose of insulin intake, and carbohydrate intake (Carbs). A two-step data preprocessing approach is employed to make the dataset fit for the training. First, the zero-values for the 'Carbs' column are replaced with the average carbohydrate value specific to each corresponding patient.

Also, the dataset was observed to have no values in the target columns at some points. Such rows are also removed before using the data for training. The preprocessing step results in a cleaner and standardised dataset, thereby laying a strong foundation for the training and prediction.

C. Simple Recurrent Neural Network (RNN)

An RNN, or Recurrent Neural Network, is category of artificial neural network which can efficiently deal with sequential or time-dependent data. RNN comprise of input layers, hidden layers, and an output layer. RNNs are famous for their unique architecture, where the output of each layer serves as input for the succeeding layer. Unlike other traditional models considering self-determining inputs, RNNs integrate output from previous steps, which gives it the ability to capture temporal dependencies within sequences. The hidden states within RNNs are for memorizing the critical details about the sequence. In essence, an RNN can be envisioned as a succession of neural networks, trainable sequentially through backpropagation, and it excels at handling data with inherent temporal relationships [17]. The basic structure of a simple RNN is provided in Figure 1.

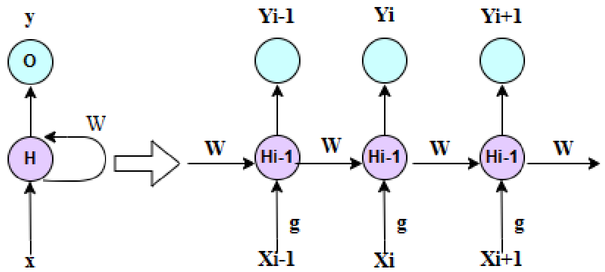

Fig 1: Structure of Simple RNN

In RNN an input layer may be followed by series of hidden layers, which are responsible for weight assignment, and for learning the complex characteristics of the data. Each layer has a bias based on the result from its preceding layer. This typical structure of execution promotes the network's ability to learn and adapt to the intricacies of the sequential data it processes [2].

D. Gated Recurrent Unit (GRU)

The Gated Recurrent Unit is another important neural network among the recurrent neural network's family. It is brilliant in way that allows only the necessary information for each time step. GRU contains two additional gates compared to the other models: the reset gate and the update gate. The reset gate controls how much of the information from the previous stage must be forgotten, adding an element of alteration to the number. The update gate is responsible for the incorporation of new input into the hidden state, regulating the extent to which the network should embrace fresh information. This interaction of gates enables GRU to calculate its output based on the updated hidden state, offering a nuanced approach to capturing temporal dependencies in sequential data [4].

The Gated Recurrent Unit (GRU) has only three gates The knowledge preserved in the Internal Cell State of an LSTM unit becomes part of the hidden state of the GRU. The distinct gates within a GRU, including the Update Gate (z) and the Reset Gate (r), play analogous roles to the Output Gate and an amalgamation of the Input Gate and Forget Gate in an LSTM, respectively. Furthermore, the often-ignored current Memory Gate adds non-linearity and zero-mean properties to the input, strategically diminishing the role of previously received information on the present data sent to the future. Such a complex structure gives GRU an outstanding level of freedom and effectiveness at capturing complex patterns in sequential data. The basic structure of a GRU cell is depicted in Figure 2.

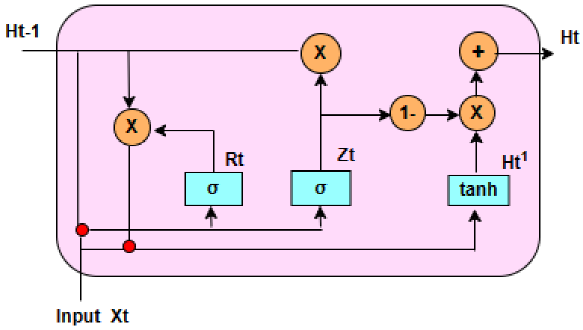

Figure 2: Structure of GRU

E. Long-Short-Term Memory (LSTM)

Long-short-term memory (LSTM) is a type of RNN used to process sequential or time-series data. They are able to discard or retain information selectively by using internal memory cells and necessary gates to control flow of information. Thus, it eliminates the vanishing gradient problem in traditional RNNs. Figure 3 illustrates the structure of a single LSTM cell.

An LSTM cell can be viewed as a series of four neural networks with memory cells connected. An LSTM unit consists of the following components: a cell, an input, an output gate, and a forget gate. The input gate is responsible for looking through the inputs and deciding on a value that should be written in or thrown away [18]. By incorporating the sigmoid function along with the inputs from the previous state and the current input, the forget gate produces a value ranging from 0 to 1. The value will decide if the information in a specific cell state may be preserved or excluded. This concept demonstrates the intricate decision-making mechanisms integrated into the LSTM architecture [5].

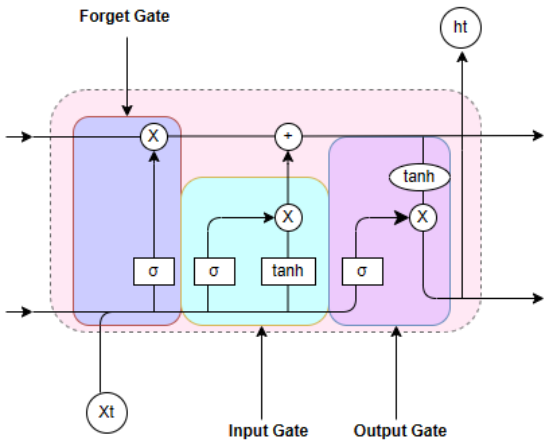

Figure 3: Structure of LSTM Cell

F. Monte Carlo Simulation

Monte Carlo Simulation is a computational method used for modelling and analysing complex systems through random sampling. This technique generates numerous random inputs within predefined ranges. Subsequently, the behaviour of the system is simulated to derive a range of probable outcomes. This process is repeated thousands or even millions of times, to reach a decision with a probabilistic perspective on various scenarios, particularly beneficial when navigating uncertainties. [6]. The fundamental idea behind the Monte Carlo simulation is to estimate the outcome by calculating the ratio of successful events to the total number of trials, adjusted by the number of possible outcomes in each trial. As the number of trials increases, the estimated outcome converges into a more precise depiction of the underlying system's behaviour.

In the context of hypoglycemia prediction, we can use it to simulate possible future blood glucose levels based on various uncertain factors.

Let's denote:

- *BGt* -> blood glucose level at time *t*,
- *It*, -> the insulin dosage at time *t*,
- *Ct*, -> the carbohydrate intake at time *t*,
- *Ut* -> other relevant factors affecting blood glucose level

The blood glucose level at the next time step (*t*+1) can be simulated using a Monte Carlo simulation step might be:

$$BGt+1=f(BGt,It,Ct,Ut)+\epsilon t \qquad [19]$$

where *f* is the function representing the relationship between the inputs and the blood glucose level, and ϵt is a random error term introduced to account for uncertainty or variability.

In the case of blood glucose prediction, Monte Carlo simulation sample random values for the input features such as CGM, insulin, Carbs etc from their corresponding probability distribution. After conducting many iterations, the results are aggregated by calculating the mean prediction of all the simulations.

This process allows for the consideration of uncertainties and variability in the factors affecting blood glucose, providing a probabilistic prediction rather than a deterministic one. The result is a range of potential outcomes, helping in understanding the likelihood of hypoglycemia under different scenarios and informing decision-making in managing diabetes.

G. Model Construction

This study concentrates on analysing the performance of Recurrent Neural Network (RNN) architectures, specifically Simple RNN, GRU, and LSTM networks in hypoglycemia prediction in Type-1 diabetic patients. As hypoglycemia, is considered as a dynamic and time-sensitive phenomenon RNN architectures are well-suited to analyze the time-series nature of Continuous Glucose Monitoring (CGM) datasets. By leveraging the sequential nature of the data, RNN models contribute to the development of predictive tools that can anticipate hypoglycemic events within specific time frames, ultimately enhancing the proactive management of T1DM.

The primary objective of our research is to develop robust models that can effectively predict hypoglycemic events within a timeframe of 60, 75, and 90 minutes (denoted as 'Hypo60', ‘Hypo75', and ‘Hypo90') based on features such as age, CGM measurements, insulin doses, and carbohydrate intake.

The model construction begins with the definition of a generic RNN architecture which takes a cell type (Simple RNN, GRU, or LSTM) as an argument, creating a Sequential model with a single layer of the specified cell type, followed by a Dense layer with sigmoid activation. The input shape is set to (4, 1), representing the four features (Age, CGM, Insulin, Carbs) at each time step.

The dataset is then pre-processed and reshaped for RNN input. The training set, consisting of features and corresponding target variable 'Hypo60', is reshaped into a three-dimensional array suitable for RNN input. The process is repeated for the testing set, ensuring consistency in model evaluation.

Monte Carlo simulations are conducted to account for the inherent variability in model training. For each simulation, the defined RNN models (Simple RNN, GRU, and LSTM) are instantiated and trained on the reshaped training data the number of epochs for a set (in this case, 10 epochs). The trained models are then used to make predictions on the testing set. The results from each simulation are stored in separate lists, allowing for subsequent analysis of the predictive performance across multiple runs. The experiment is then repeated for prediction horizons of 75 minutes and 90 minutes.

This methodology ensures a comprehensive evaluation of the RNN models, considering the inherent stochasticity in training neural networks. A block diagram representing the methodology is given in Figure 4. A detailed algorithm for the described model is given below.

Algorithm:

```
1. Data Loading:
Load the dataset:
     D = LoadData("diabetes_data.xlsx")

2. Data Preprocessing:
a. For each patient p:
   For each row where 'Carbs' is zero:
     Replace carbs by the mean of the carbs of the same patient
b. Remove rows where target values are missing
3. Data Splitting
Separate training and Testing data:
  D_train, D_test = TrainTestSplit(D, test_size=0.2,
random_state=42)
4. Model Definition:
State the model creation function:
function create_rnn_model(cell_type):
    M = RNNModel(cell_type, num_cells=64)
    Add one Dense layer with
     activation='sigmoid'))
    compile with loss='binary_crossentropy', optimizer='adam',
metrics=['accuracy'])
    return M
5. Data Reshaping: Reshape data for RNN input:
   X_train, y_train = ReshapeForRNN(D_train)
   X_test, y_test = ReshapeForRNN(D_test)

6. Monte Carlo Simulations:
Number of simulations: 8

For s in range(num_simulations):
```

```
    Instantiate and train models:
    M_rnn = create_rnn_model(SimpleRNN)
    M_gru = create_rnn_model(GRU)
    M_lstm = create_rnn_model(LSTM)

    TrainModel(M_rnn, X_train, y_train, epochs=10, batch_size=32)
    TrainModel(M_gru, X_train, y_train, epochs=10, batch_size=32)
    TrainModel(M_lstm, X_train, y_train, epochs=10, batch_size=32)
```

7. Prediction Aggregation: Initialize lists to store predictions:

```
y_pred_rnn_list = [], y_pred_gru_list = [], y_pred_lstm_list = []
For s in range(num_simulations):
    Append predictions for each model:
y_pred_rnn_list.append(MakePredictions(M_rnn, X_test))
y_pred_gru_list.append(MakePredictions(M_gru, X_test))
y_pred_lstm_list.append(MakePredictions(M_lstm, X_test))
```

8. Performance Metrics Calculation:

a. Calculate mean predictions:

```
  y_pred_rnn_mean = Mean(y_pred_rnn_list)
  y_pred_gru_mean = Mean(y_pred_gru_list)
  y_pred_lstm_mean = Mean(y_pred_lstm_list)
```

b.Calculate performance metrics for each model:

```
  metrics_rnn = CalculateMetrics(y_test, y_pred_rnn_mean)
  metrics_gru = CalculateMetrics(y_test, y_pred_gru_mean)
  metrics_lstm = CalculateMetrics(y_test, y_pred_lstm_mean)
```

9. Confusion Matrix Calculation: Calculate confusion matrices:

```
confusion_matrix_rnn = ConfusionMatrix(y_test, y_pred_rnn_mean)
confusion_matrix_gru = ConfusionMatrix(y_test, y_pred_gru_mean)
confusion_matrix_lstm = ConfusionMatrix(y_test, y_pred_lstm_mean)
```

10. Repeat steps 1 to 9 for target variables ‘Hypo 75’ and ‘Hypo90’.

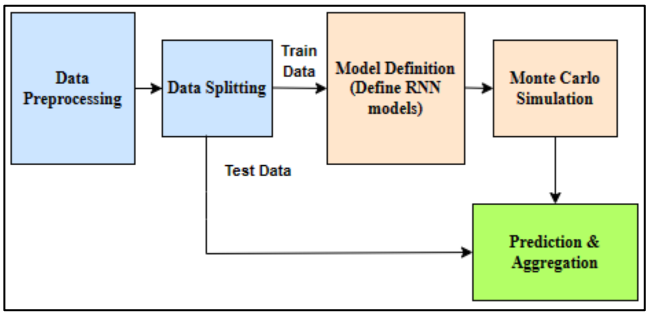

Figure 4: Methodology Diagram of the prescribed model

3. Result and Discussion

The prescribed model is a performance analysis of Recurrent Neural Networks (RNNs) for the prediction of hypoglycemia events in diabetic patients - specifically Simple RNN, GRU (Gated Recurrent Unit), and LSTM (Long Short-Term Memory) models. In order to evaluate the performance of the RNN models for hypoglycemia prediction, several key metrics are considered: accuracy, sensitivity, specificity, precision, and the F1 score.

Accuracy is the ability of the model to predict the two cases of hypoglycemia and non hypoglycemia correctly. Sensitivity assesses the model's ability to identify hypoglycemia cases among actual positives, reducing false negatives. Specificity measures the model's proficiency in identifying cases without hypoglycemia, Precision focuses on the reliability of positive predictions by calculating the proportion of properly predicted positives to total predicted positives. Lastly, the F1 score assess the balance precision and sensitivity, especially valuable in addressing imbalances in positive and negative instances [7].

In this comprehensive experiment, the predictive capabilities of Simple RNN, GRU, and LSTM models were rigorously assessed in the context of forecasting hypoglycemia. The models were input with age, carbs intake, insulin dosage, and previous CGM readings. Their performance was meticulously scrutinized across distinct prediction horizons, specifically set at 60, 75, and 90 minutes. We shall consider the performance of each model one by one.

A. Simple RNN

The Simple RNN model exhibits varying degrees of success in predicting hypoglycemia across different prediction horizons (PH60, PH75, PH90). For PH60, the model achieves an accuracy of 88.99%, showcasing proficiency in overall classification. The trade-off between sensitivity (35.58%) and specificity (97.23%) highlights challenges in capturing true positive hypoglycemic cases while maintaining high specificity. Precision and F1 Score values for PH60 (66.43% and 46.34%, respectively) indicate the model's reliability in making positive predictions, with room for improvement in achieving a balanced trade-off between precision and recall. Performance slightly diminishes for PH75 (sensitivity: 19.03%, Precision: 73.91%, F1 Score: 30.27%) and PH90 (sensitivity: 13.75%, Precision: 66.07%, F1 Score: 22.77%), indicating difficulties in correctly identifying hypoglycemic cases at extended prediction horizons. The evaluation of the prescribed model using Simple RNN is summarised in Table 2.

Evaluation Metric	PH60	PH75	PH90
Accuracy	0.8914	0.8824	0.8744
Specificity	0.9723	0.9896	0.989
Sensitivity	0.3558	0.1903	0.1375

Precision	0.6643	0.7391	0.6607
F1Score	0.4634	0.3027	0.2277

Table 2: Performance evaluation of the Simple RNN model

The confusion matrix for the Simple RNN model for the three prediction horizons is shown below.

Confusion Matrix for PH60:

[[1684 48]
[172 95]]

Confusion Matrix for PH75:

[[1713 18]
[217 51]]

Confusion Matrix for PH90:

[[1711 19]
[232 37]]

B. GRU

For GRU across different prediction horizons (PH60, PH75, PH90), the model demonstrates reasonable accuracy, with values ranging from 87.79% to 88.99%. Table 3 describes the performance evaluation of the model based on GRU for the varying prediction horizons.

Performance Metric	PH60	PH75	PH90
Accuracy	0.8899	0.8799	0.8779
Specificity	0.9677	0.9804	0.9884
Sensitivity	0.3858	0.2313	0.1673
Precision	0.6478	0.6458	0.6923
F1Score	0.4836	0.3407	0.2695

Table 3: Performance evaluation of the GRU model

The specificity remains consistently high, indicating the model's proficiency in correctly identifying non-hypoglycemic instances. However, there is a notable trade-off in sensitivity, especially for extended prediction horizons (PH75 and PH90), suggesting challenges in correctly capturing hypoglycemic cases. Precision and F1 Score values for all horizons emphasize the model's reliability in positive predictions, but there is room for improvement in achieving a balanced trade-off between precision and sensitivity. These results suggest that GRU performs reasonably well in predicting hypoglycemia, particularly at shorter prediction horizons.

The confusion matrix for the GRU-based model for the time frames of 60, 75, and 90 minutes is shown below.

Confusion Matrix for PH60:

[[1676 56]

[164 103]]

Confusion Matrix for PH75:

[[1697 34]

[206 62]]

Confusion Matrix for PH90:

[[1710 20]

[224 45]]

C. LSTM

In the case of LSTM for different prediction horizons (PH60, PH75, PH90), the model exhibits consistent accuracy, ranging from 87.79% to 88.94%. The overall performance of the LSTM-based model is visualized in Table 4.

Performance Metric	PH60	PH75	PH90
Accuracy	0.8894	0.8824	0.8779
Specificity	0.974	0.9752	0.9896
Sensitivity	0.3408	0.2836	0.1599
Precision	0.6691	0.6387	0.7049
F1Score	0.4516	0.3928	0.2606

Table 4: Performance evaluation of the LSTM model

Notably, the specificity remains high, indicating the model's proficiency in accurately identifying instances without hypoglycemia. However, a discernible trade-off is observed in sensitivity, especially for extended prediction horizons (PH75 and PH90), suggesting challenges in effectively capturing hypoglycemic cases. Precision and F1 Score values highlight the model's reliability in positive predictions, yet there is room for improvement in achieving a more balanced trade-off between precision and sensitivity.

The confusion matrix for the LSTM-based model is provided for the time frames of 60, 75, and 90 minutes.

Confusion Matrix for PH60:

[[1687 45]

[176 91]]

Confusion Matrix for PH75:

[[1688 43]

[192 76]]

Confusion Matrix for PH90:

[[1712 18]

[226 43]]

D. Result Analysis

In comparing the performances of the three models for the prediction horizon of 60 minutes (PH60), all models—Simple RNN, GRU, and LSTM—demonstrate robust accuracy, with Simple RNN and GRU both achieving an accuracy of 88.99%, and LSTM close following at 88.94%. Regarding specificity, Simple RNN, and LSTM exhibit similarly, high values of 97.23% and 97.40%, respectively, while GRU slightly trails at 96.77%. For sensitivity, GRU outperforms the others with a value of 38.58%, whereas Simple RNN and LSTM have sensitivities of 35.58% and 34.08%, respectively. In terms of precision, LSTM leads with 66.91%, followed closely by Simple RNN at 66.43%, and GRU at 64.78%. The F1 Score, representing a harmonic mean of precision and sensitivity, is highest for GRU at 48.36%, followed by Simple RNN at 46.34%, and LSTM at 45.16%. A graphical representation of the same is provided in Figure 5.

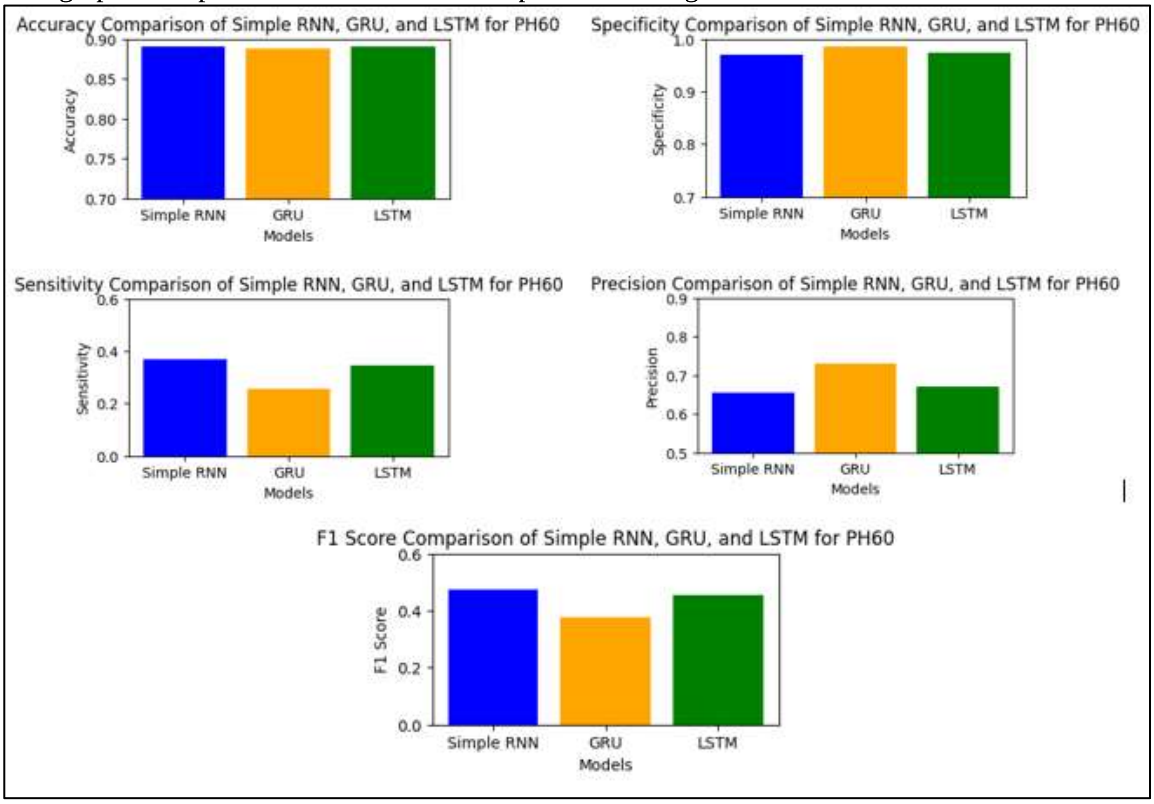

Figure 5: Performance Evaluation of Simple RNN, GRU, and LSTM for PH60

For the prediction horizon of 75 minutes (PH75), comparing the performances of the three models reveals nuanced differences. In terms of accuracy, all models—Simple RNN, GRU, and LSTM—display similar performance, with values ranging from 87.99% to 88.24%. Notably, specificity values are consistently high, indicating proficiency in correctly identifying instances without hypoglycemia. Simple RNN achieves the highest specificity at 98.96%, followed closely by LSTM at 97.52%, and GRU at 98.04%. However, the trade-off between sensitivity and specificity varies, with GRU leading in sensitivity at 23.13%, followed by LSTM at 28.36%, and Simple RNN at 19.03%. Precision values show a similar trend, with Simple RNN having the highest precision at 73.91%, followed by LSTM at 63.87%, and GRU at 64.58%. The F1 Score, reflecting a balance between precision and sensitivity, is highest for LSTM at 39.28%, followed by GRU at 34.07%, and Simple RNN at 30.27%. These results indicate that for PH75, LSTM demonstrates a slightly superior balance between sensitivity and precision compared to the other models. Figure 6 represents the performance evaluation of Simple RNN, GRU, and LSTM models for PH75.

For the prediction horizon of 90 minutes (PH90), a comparative analysis of the three models—Simple RNN, GRU, and LSTM—reveals similar overall accuracies, ranging from 87.44% to 87.79%. Specificity values remain consistently high across all models, with Simple RNN and GRU achieving 98.90% and 98.84%, respectively, and LSTM closely following at 98.96%. However, sensitivity values are lower for this extended prediction horizon, with GRU leading at 16.73%, followed by LSTM at

15.99%, and Simple RNN at 13.75%. Precision values vary, with LSTM demonstrating the highest precision at 70.49%, followed by GRU at 69.23%, and Simple RNN at 66.07%. The F1 Score, capturing the balance between precision and sensitivity, is highest for LSTM at 26.06%, followed by GRU at 26.95%, and Simple RNN at 22.77%. These findings suggest that for PH90, the models exhibit comparable performance with slight variations in sensitivity and precision, emphasizing the challenges associated with predicting hypoglycemia over longer timeframes. A graphical representation of the performance of the three models is provided in Figure 7. Table 5 illustrates a performance comparison of the three models throughout the timeframes.

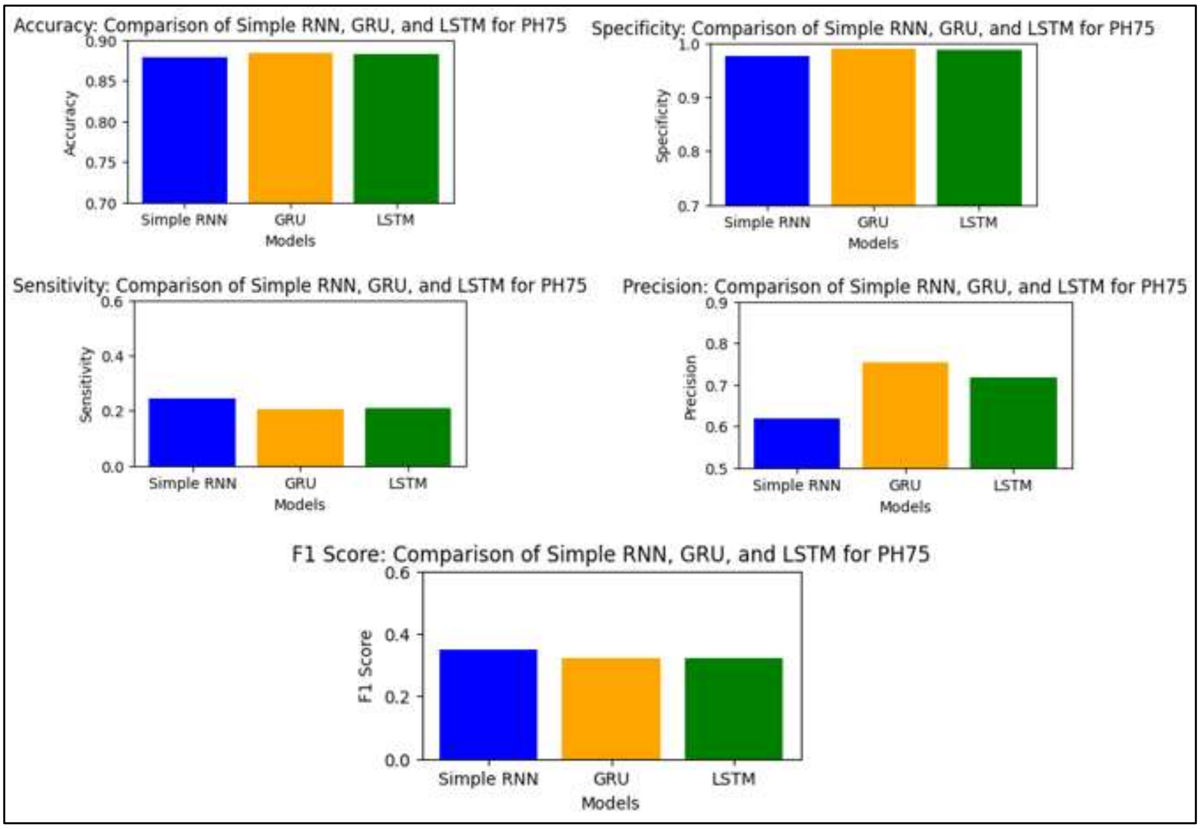

Figure 6: Performance Evaluation of Simple RNN, GRU, and LSTM for PH75

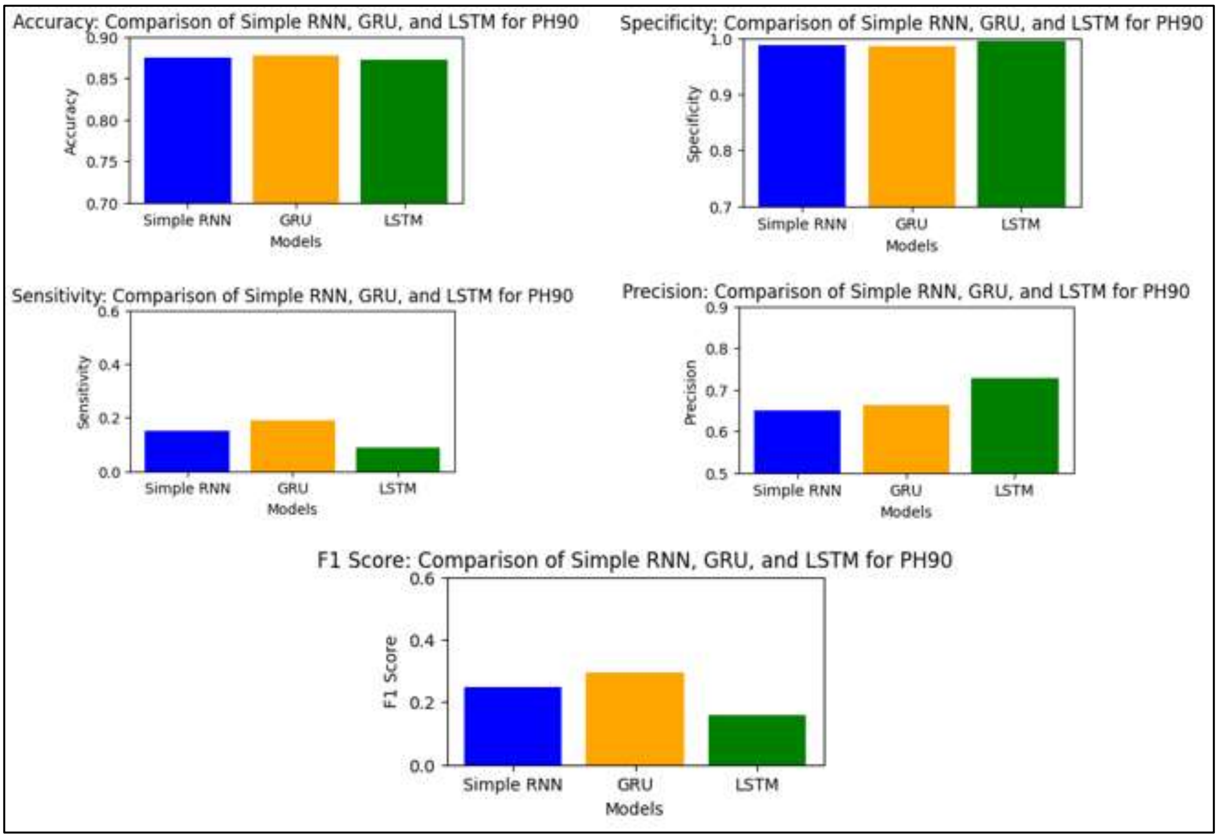

Figure 7: Performance Evaluation of Simple RNN, GRU, and LSTM for PH90

Table 5: Comparison chart of Overall performance of the models

Metrics	Models	PH60	PH75	PH90
Accuracy	Simple RNN	0.8914	0.8824	0.8744
	GRU	0.8899	0.8799	0.8779
	LSTM	0.8894	0.8824	0.8779
Specificity	Simple RNN	0.9723	0.9896	0.989
	GRU	0.9677	0.9804	0.9884
	LSTM	0.974	0.9752	0.9896
Sensitivity	Simple RNN	0.3558	0.1903	0.1375
	GRU	0.3858	0.2313	0.1673
	LSTM	0.3408	0.2836	0.1599
Precision	Simple RNN	0.6643	0.7391	0.6607
	GRU	0.6478	0.6458	0.6923
	LSTM	0.6691	0.6387	0.7049
F1Score	Simple RNN	0.4634	0.3027	0.2277
	GRU	0.4836	0.3407	0.2695
	LSTM	0.4516	0.3928	0.2606

Based on the study encompassing the specified timeframes, each model—Simple RNN, GRU, and LSTM—demonstrates its strengths and trade-offs. For PH60, all models perform comparably, showcasing high accuracy and specificity, with subtle differences in sensitivity and precision. For

PH75, LSTM exhibits a slightly superior balance between sensitivity and precision. However, for the extended prediction horizon of PH90, all models face challenges with lower sensitivity values. Considering the trade-offs and overall performance, LSTM emerges as a promising choice, consistently maintaining high specificity and demonstrating a balanced trade-off between sensitivity and precision.

4. Conclusion and future scope

The study thoroughly analysed and compared the predictive capabilities of Simple RNN, GRU, and LSTM models across prediction horizons (PH60, PH75, PH90) for hypoglycemia prediction. The models showed commendable accuracies, with nuanced differences in specificity, sensitivity, precision, and F1 Score values. When the balance between sensitivity and precision is considered, GRU shoed higher F1 Score for PH 60 and LSTM is observed to be the best for PH75, while all models faced challenges in maintaining sensitivity for the extended PH90. The study suggests LSTM as a promising choice for its consistent performance, showing higher accuracy and F1 score. But the model can be further refined and should be validated against diverse datasets to enhance robustness and generalizability.

Reference

[1] Zhao, Q., Zhu, J., Shen, X. et al. Chinese diabetes datasets for data-driven machine learning. Sci Data 10, 35 (2023). https://doi.org/10.1038/s41597-023-01940-7

[2] Zhu T, Li K, Chen J, Herrero P, Georgiou P. Dilated Recurrent Neural Networks for Glucose Forecasting in Type 1 Diabetes. J Healthc Inform Res. 2020 Apr 12;4(3):308-324. doi: 10.1007/s41666-020-00068-2. PMID: 35415447; PMCID: PMC8982716.

[3] A. D. Association. Standards of medical care in diabetes – 2017. Diabetes Care, vol. 40, no. 1, pp. S1–S135, 2018

[4] Kim DY, Choi DS, Kim J, Chun SW, Gil HW, Cho NJ, Kang AR, Woo J. Developing an Individual Glucose Prediction Model Using Recurrent Neural Network. Sensors (Basel). 2020 Nov 12;20(22):6460. doi: 10.3390/s20226460. PMID: 33198170; PMCID: PMC7696446.

[5] Mirshekarian S, Shen H, Bunescu R, Marling C. LSTMs and Neural Attention Models for Blood Glucose Prediction: Comparative Experiments on Real and Synthetic Data. Annu Int Conf IEEE Eng Med Biol Soc. 2019 Jul;2019:706-712. doi: 10.1109/EMBC.2019.8856940. PMID: 31945995; PMCID: PMC7890945.

[6] Adekitan, A.I. (2014). MONTE CARLO SIMULATION. 10.13140/RG.2.2.15207.16806.

[7] Chicco, D., Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21, 6 (2020). https://doi.org/10.1186/s12864-019-6413-7

[8] Vehí J, Contreras I, Oviedo S, Biagi L, Bertachi A. Prediction and prevention of hypoglycaemic events in type-1 diabetic patients using machine learning. Health Informatics Journal. 2020;26(1):703-718. doi:10.1177/1460458219850682

[9] Kim DY, Choi DS, Kim J, Chun SW, Gil HW, Cho NJ, Kang AR, Woo J. Developing an Individual Glucose Prediction Model Using Recurrent Neural Network. Sensors (Basel). 2020 Nov 12;20(22):6460. doi: 10.3390/s20226460. PMID: 33198170; PMCID: PMC7696446.

[10] Ganjar Alfian, Muhammad Syafrudin, Muhammad Anshari, Filip Benes, Fransiskus Tatas Dwi Atmaji, Imam Fahrurrozi, Ahmad Fathan Hidayatullah, Jongtae Rhee, Blood glucose prediction model for type 1 diabetes based on artificial neural network with time-domain features, Biocybernetics and Biomedical Engineering, Volume 40, Issue 4, 2020, Pages 1586-1599, ISSN 0208-5216, https://doi.org/10.1016/j.bbe.2020.10.004.

[11] Ning Ma, Xia Yu, Tao Yang, Yuhang Zhao, Hongru Li, A hypoglycemia early alarm method for patients with type 1 diabetes based on multi-dimensional sequential pattern mining, Heliyon, Volume 8, Issue 11, 2022, e11372, ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2022.e11372.

[12] Tsichlaki S, Koumakis L, Tsiknakis M. Type 1 Diabetes Hypoglycemia Prediction Algorithms: Systematic Review. JMIR Diabetes. 2022 Jul 21;7(3):e34699. doi: 10.2196/34699. PMID: 35862181; PMCID: PMC9353679.

[13] Zhu, T., Uduku, C., Li, K. et al. Enhancing self-management in type 1 diabetes with wearables and deep learning. npj Digit. Med. 5, 78 (2022). https://doi.org/10.1038/s41746-022-00626-5

[14] Peng X, Li Q, Wang Y, Yan D (2022) High-precision blood glucose prediction and hypoglycemia warning based on the LSTM-GRU model. Trends Comput Sci Inf Technol 7(3): 074-080. DOI: 10.17352/tcsit.000053

[15] Thomsen HB, Jakobsen MM, Hecht-Pedersen N, Jensen MH, Kronborg T. Prediction of Hypoglycemia From Continuous Glucose Monitoring in Insulin-Treated Patients With Type 2 Diabetes Using Transfer Learning on Type 1 Diabetes Data: A Deep Transfer Learning Approach. Journal of Diabetes Science and Technology. 2023;0(0). doi:10.1177/19322968231215324

[16] Zhu, T., Li, K., Chen, J. et al. Dilated Recurrent Neural Networks for Glucose Forecasting in Type 1 Diabetes. J Healthc Inform Res 4, 308–324 (2020). https://doi.org/10.1007/s41666-020-00068-2

[17] Fayrouz Allam, Zaki Nossai, Hesham Gomma, Ibrahim Ibrahim, Mona Abdelsalam. A Recurrent Neural Network Approach for Predicting Glucose Concentration in Type-1 Diabetic Patients. 12th Engineering Applications of Neural Networks (EANN 2011) and 7th Artificial Intelligence Applications and Innovations (AIAI), Sep 2011, Corfu, Greece. pp.254-259, ff10.1007/978-3-642-23957-1_29ff. ffhal01571345.

[18] Yong Yu, Xiaosheng Si, Changhua Hu, Jianxun Zhang; A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. Neural Comput 2019; 31 (7): 1235–1270. doi: https://doi.org/10.1162/neco_a_01199

[19] Harrison RL. Introduction To Monte Carlo Simulation. AIP Conf Proc. 2010 Jan 5;1204:17-21. doi: 10.1063/1.3295638. PMID: 20733932; PMCID: PMC2924739.