pHealth 2024
M. Giacomini et al. (Eds.)

doi:10.3233/SHTI240067

GLSTM: On Using LSTM for Glucose Level Prediction

Muhammad KASHIF[a], Sergio FLESCA[a,1] and Pierangelo VELTRI[a]
[a] *DIMES University of Calabria, Italy*
ORCiD ID: Muhammad Kashif https://orcid.org/0000-0002-5170-8004
ORCiD ID: Prof. Sergio Flesca https://orcid.org/0000-0002-4164-940X
ORCiD ID: Prof. Pierangelo Veltri https://orcid.org/0000-0003-2494-0294

Abstract. The Prediabetes impacts one in every three individuals, with a 10% annual probability of transitioning to type 2 diabetes without lifestyle changes or medical interventions. It's crucial to manage glycemic health to deter the progression to type 2 diabetes. In the United States, 13% of individuals (18 years of age and older) have diabetes, while 34.5% meet the criteria for prediabetes. Diabetes mellitus and prediabetes are more common in older persons. Currently, nevertheless, there aren't many noninvasive, commercially accessible methods for tracking glycemic status to help with prediabetes self-management. This study tackles the task of forecasting glucose levels using personalized prediabetes data through the utilization of the Long Short-Term Memory (LSTM) model. Continuous monitoring of interstitial glucose levels, heart rate measurements, and dietary records spanning a week were collected for analysis. The efficacy of the proposed model has been assessed using evaluation metrics including Root Mean Square Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), and the coefficient of determination (R2).

Keywords. Glucose prediction, deep learning, LSTM, prediabetes, diabetes, wearable devices

1. Introduction

Prediabetes, with elevated but not diabetic-level blood glucose, often progresses to diabetes. Diabetes is a major global health concern, with rising prevalence, necessitating regular blood glucose monitoring to prevent severe complications like cardiac issues, strokes, and kidney damage [1] [2]. Blood glucose definitions distinguish between Type 1 Diabetes (T1D) and Type 2 Diabetes (T2D). T1D involves autoimmune destruction of insulin-producing cells, while T2D stems from lifestyle and genetic factors causing insulin resistance or insufficient production [3]. T2D constitutes 90-95% of cases and includes hyperglycemia and hypoglycemia. Managing blood glucose levels involves addressing dietary choices, medication, and physical activity for effective diabetes control [4].

[1] Corresponding Author: Prof. Sergio Flesca, DIMES - UNICAL flesca@dimes.unical.it

In the US, 37.3 million adults have diabetes, mainly Type 2 (T2D), with 96 million having prediabetes [5]. Diabetes costs $327 billion annually, with healthcare costs 2.3 times higher for diabetics. It's a leading non-communicable disease, and global projections predict a rise to 578 million cases by 2030 and 700 million by 2045 [6]. Digital health tech, including smartphones and wearables, offers personalized interventions for disease management and prevention [7]. Wearable devices with biosensors provide seamless monitoring of health parameters like skin temperature and heart rate, integrating into daily life [8].

The main contribution of this study is to introduce an LSTM deep learning approach for predicting blood glucose levels on a personalized prediabetes participant dataset [9]. Pre-processing of sensor data is conducted, which involves ensuring sampling consistency and interpolating to address missing samples. Furthermore, the effectiveness of the proposed model is assessed using metrics such as RMSE, MSE, MAE, and R2, and its performance is compared with state-of-the-art methods.

2. Data Collection and Methods

Following the introduction to prediabetes and diabetes, this section elaborates on the dataset, data preprocessing, and proposed methodology in detail.

2.1. Data Collection and Preprocessing

Data from 16 participants monitored normal blood glucose levels using Dexcom G6 and Empatica E4 devices for 8-10 days each, excluding those with certain medical histories or lacking access to necessary technology. Data for each participant is stored in folders numbered 001 to 016, with separate CSV files containing timestamped values for seven features including glucose, physiological signals, and food intake details, alongside demographic information [4]. The prediabetes dataset[2] is publicly available [9].

The dataset underwent rigorous preprocessing, excluding entries from participants 003, 007, 013, 015, and 016 due to irregularities such as data recorded in different months, which could impact performance. After this pruning phase, the dataset contains data of 11 prediabetics patients. This data is characterized by the following features: glucose levels, heart rate measurements, sugar intake, and carbohydrate consumption, measured over one week for each patient, as reported in Figure 1. A linear approximation method synchronized these features to ensure dataset coherence, facilitating meaningful analysis.

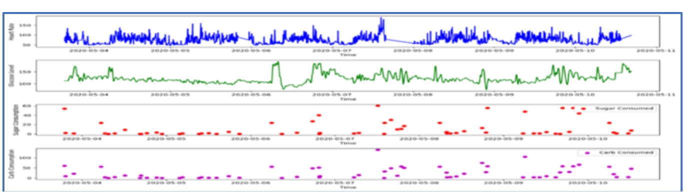

Figure 1: Different features are shown from different sensors and food log for 014 participant such as heart rate, glucose level, sugar, and carb.

[2] BIG IDEAs Lab Glycemic Variability and Wearable Device Data v1.1.2 (physionet.org)

2.2. Glucose Prediction Model

LSTM, a type of recurrent neural network (RNN), excels at processing diverse data types like speech and images. It's widely used in deep learning for tasks such as speech and handwriting recognition. LSTM's unique design includes gates to handle long-term dependencies effectively. This makes it suitable for complex tasks like glucose prediction with extended observation windows [11]. The TensorFlow LSTM Sequential model is designed for sequence prediction tasks, comprising an input layer, LSTM layer, and dense layer. The input layer receives input sequence features, while the LSTM layer learns temporal dependencies with its memory cells and gates. The final prediction is generated by the fully connected Dense layer, as illustrated in the LSTM flow Figure 2.

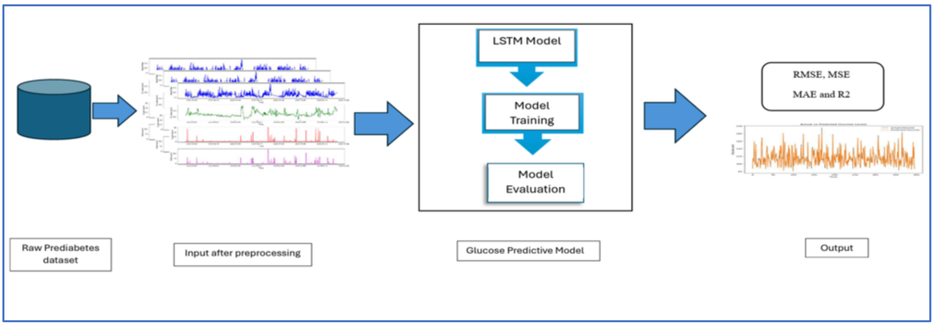

Figure 2: Architecture of the proposed model

3. Experimental Analysis and Discussion

We utilized the Python programming language along with various libraries to preprocess the data and develop predictive models. First necessary libraries including pandas, NumPy, scikit-learn, and TensorFlow were imported for data manipulation, numerical computations, and deep learning tasks, respectively. Followed by outline steps to our approach sequences of glucose values along with related features patient were created. standard scaler is used to scale the data followed by splitting it into training and testing sets. As input sequences to the predictive model 36 lookback values means 3-hour horizon were iteratively used for over 7 days for each participant. LSTM-based deep learning model using TensorFlow's Keras API is deployed by training the prediabetes data. The performance of trained model is evaluated based on various metrics such as MSE, RMSE, MAE, and R2. Hyperparameters used for the proposed model are presented in Table 1.

Table 1: Hyperparameters

Hyperparameter	Value
Lookback	36
Batch size	32
Activation function	relu
Optimization algorithm	Adam
Loss function	MSE
Number of epochs	150

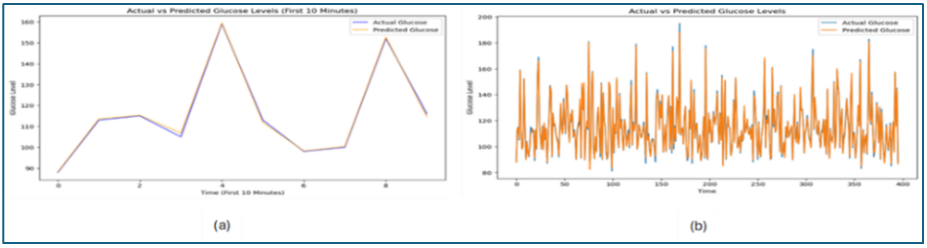

Figure 3: The comparison between actual and predicted glucose levels for participant 014 is illustrated. (a) displays the first 10-minute interval, while (b) presents the data for all intervals.

Extensive analysis has been conducted and the comparison of actual vs predicted glucose levels has been illustrated in Figure 3. Two sets of experiments were undertaken to evaluate the performance of the LSTM model in predicting glucose values. Initially, the LSTM model was assessed using only the historical glucose feature. The proposed model exhibited better performance as evidenced by improved metrics including RMSE, MSE, MAE, and the coefficient of determination (R2), achieving average values of 3.70, 15.08,1.82 and 0.97, respectively.

Table 2: Comparison with state of the art

Reference	Technique	RMSE
[11]	LSTM	4.906
[10]	LSTM	5.04
Proposed Model Experiment I	**LSTM**	**3.70**
Proposed Model Experiment II	**LSTM**	**1.83**

Subsequently, an expanded feature set comprising additional physiological parameters such as heart rate, along with glucose levels and dietary information including sugar and carbohydrate intake, was incorporated into the LSTM model for glucose prediction. The augmented model demonstrated enhanced predictive capabilities, as indicated by improved RMSE, MSE, MAE, and R2 scores, achieving average values of 1.83, 3.44, 1.24 and 0.99, respectively. The performance comparison of the proposed model is described in Table 2. LSTM outperforms ARIMA for blood glucose prediction, with MSE and RMSE of 24.075 and 4.906, respectively, compared to 108.140 and 10.399 for ARIMA [11]. An LSTM approach accurately predicts glucose levels, with MSE of 5.04 and MARD of 2.61, mitigating hypoglycemia risks due to insulin dosage delays [10].

Summarizing, the fact that the model used for experiment II yield a better accuracy compared with the model used for experiment I suggests that the additional features used for yielding predictions in experiments II have a relevant impact on the variations of the glucose levels, opening the way for further analysis of this aspect.

4. Conclusions and Future Work

This study introduces an efficient approach for precise and continuous blood glucose level prediction in individuals with prediabetes. Data preprocessing, including interpolation techniques are evaluated to enhance data quality. A Long Short-Term

Memory (LSTM) model based on recurrent neural networks is trained for glucose prediction, outperforming existing tools. However, the current study is limited by its application at prediabetics patients only. The approach needs to be tested over diabetics patients data in order to prove its utility in practical contexts. Nonetheless, the good performances exhibited by the LSTM model encourage its application in context where additional digital biomarkers are available where higher accuracy levels may be achieved, particularly by integrating features like accelerometry and skin temperature.

References

[1] Skoglund G, Nilsson BB, Olsen CF, Bergland A, Hilde G. Facilitators and barriers for lifestyle change in people with prediabetes: a meta-synthesis of qualitative studies. BMC Public Health. 2022; 22(1). doi:10.1186/s12889-022-12885-8

[2] Marik M, editor. Machine Learning Based Continuous Glucose Monitoring System. Vol. 6. Global Journal on Innovation, Opportunities and Challenges in AAI and Machine Learning; 2022.

[3] Patro KK, Allam JP, Sanapala U, Marpu CK, Samee NA, Alabdulhafith M, et al. An effective correlation-based data modeling framework for automatic diabetes prediction using machine and deep learning techniques. BMC Bioinformatics 2023; 24(1). Doi:10.1186/s12859-023-05488-6

[4] Bent B, Cho PJ, Henriquez M, Wittmann A, Thacker C, Feinglos M, et al. Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches. NPJ Digit Med. 2021;4(1). doi:10.1038/s41746-021-00465-w

[5] Zahedani AD, McLaughlin T, Veluvali A, Aghaeepour N, Hosseinian A, Agarwal S, et al. Digital health application integrating wearable data and behavioral patterns improves metabolic health. NPJ Digit Med. 2023;6(1):1–15.

[6] Adams D, Nsugbe E. A pilot on the endocrine effects of hormonal replacement therapy on menopausal T1 diabetics using wearable sensors. In: The 9th International Electronic Conference on Sensors and Applications. Basel Switzerland: MDPI; 2022. p. 19.

[7] van den Brink WJ, van den Broek TJ, Palmisano S, Wopereis S, de Hoogh IM. Digital biomarkers for personalized nutrition: Predicting meal moments and interstitial glucose with non-invasive, wearable technologies. Nutrients. 2022; 14(21):4465.

[8] Ahmed A, Aziz S, Qidwai U, Abd-Alrazaq A, Sheikh J. Performance of artificial intelligence models in estimating blood glucose level among diabetic patients using non-invasive wearable device data. Comput Methods Programs Biomed Update. 2023;3(100094):100094.
doi:10.1016/j.cmpbup.2023.100094

[9] Cho P, Kim J, Bent B, Dunn J. BIG IDEAs lab glycemic variability and wearable device data. PhysioNet; 2023. doi:10.13026/ZTHX-5212

[10] Kalita D, Mirza KB. Glucose prediction using wide-deep LSTM network for accurate insulin dosing in artificial pancreas. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2022. p. 4426–9.

[11] Fayaz L, Joseph R, Ankayarkanni B, PrinceMary S, Asha P. Multi-scale and context aware optimized glucose prediction using neural networks. In: 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI). IEEE; 2023. p. 820–5.

[12] Iacono F, Magni L, Toffanin C. Personalized LSTM models for glucose prediction in Type 1 diabetes subjects. In: 2022 30th Mediterranean Conference on Control and Automation (MED). IEEE; 2022. p. 324–9.

[13] Butt H, Khosa I, Iftikhar MA. Feature transformation for efficient blood glucose prediction in type 1 diabetes Mellitus patients. Diagnostics (Basel). 2023;13(3):340.

[14] Wang Z, Xiao X, Pang Y, Su W, Kikkawa T. Noninvasive, intelligent blood glucose monitoring on fingertip using dual-band fusion and LSTM-R network. IEEE Sens J. 2024;24(3):3465–76. doi:10.1109/jsen.2023.3344230

[15] Sun Y, Cano-Garcia H, Kallos E, O'Brien F, Akintonde A, Motei D-E, et al. Random forest analysis of combined millimeter-wave and near-infrared sensing for noninvasive glucose detection. IEEE Sens J. 2023;23(17):20294–309. doi:10.1109/jsen.2023.3293248

[16] Katsarou DN, Georga EI, Christou M, Tigas S, Papaloukas C, Fotiadis DI. Short term glucose prediction in patients with type 1 diabetes mellitus. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2022. p. 329–32.