An Efficient Attention-Based LSTM Framework for Blood Glucose Level Prediction

Sunny Arora[1*], Shailender Kumar[1], Pardeep Kumar[2]

[1] Department of Computer Science Engineering, Delhi Technological University, Delhi 110042, India
[2] Department of Computer Science, Warwick Manufacturing Group (WMG), University of Warwick, Coventry CV4 7AL, UK

Corresponding Author Email: sunnyarora_2k18phdco@dtu.ac.in

ABSTRACT

Predicting blood glucose is highly significant for patients with diabetes to manage their condition efficiently. Deep learning (DL) approaches have demonstrated great potential in blood glucose prediction modeling. By leveraging time-series data from continuous glucose monitoring (CGM) devices, this model can capture complicated temporal dependencies and patterns in glucose dynamics. DL based blood glucose prediction models learn from insulin dosages, past glucose readings, physical activity levels, meal intake, and other related features to forecast future blood glucose levels with maximum accuracy. Thus, this study presents an Optimal Attention-based Long Short-Term Memory for Blood Glucose Level Prediction (OALSTM BGLP) model. Firstly, it employs Min-Max scaling to normalize the input data, ensuring consistent and meaningful comparisons across different features. Additionally, the model generates time series data for multiple forecasting horizons, including 15, 30, 45, and 60-minute(m) intervals, enabling flexible and dynamic predictions to accommodate various planning and decision-making needs. Moreover, the OALSTM-BGLP technique uses the ALSTM model, which incorporates an attention mechanism to selectively concentrate on relevant information within the input sequence while capturing long-term dependencies. This attention mechanism permits the method to effectively extract salient features from the input data, enhancing its predictive capabilities. Furthermore, the model is optimized using the RMSProp optimizer, which adjusts the rate of learning dependent on the magnitude of recent gradients, facilitating efficient training and convergence. The performance evaluation of the developed technique on the OhioT1DM dataset shows its promising performance over recent state-of-the-art methods.

1. INTRODUCTION

Diabetes mellitus is a metabolic disorder that disrupts the body's ability to regulate blood glucose (BG) levels. Beta cells in the pancreas secrete insulin, an endocrine hormone that regulates glucose utilization [1]. Type 1 diabetes is characterized by the loss of these insulin-producing beta cells, leading to insulin deficiency [2]. As a result, individuals with diabetes must continuously monitor their blood glucose levels, regulate them through consistent insulin administration, and make informed decisions about their medical regimen (e.g., once daily or before each meal) to maintain normal blood sugar level [3]. The primary objective in diabetes management is to optimize insulin dosages to prevent hyper/ hypo glycemia [4]. This is difficult due to the multitude of factors, including lifestyle, mental state, stress, and physical exercise, that also affect insulin intake, nutrition, and glucose levels [4]. Despite various advancements in diabetes observation, continuous glucose monitoring (CGM) remains aggressive. CGM devices are capable of providing an individual's glycemic condition at an assumed time. A positive recognition could dramatically expand the everyday behavior of diabetes by the patients themselves [5].

To simplify and accomplish the management of type 1 diabetes, a continual growth in automated procedures is quite essential, despite the advancements obtained thus far. In this way, commerce-precise BG devices might be game-altering [6]. These devices can provide early warnings about potential glycemic actions, enabling the implementation of either automatic or non-automated preventive measures. Moreover, these mechanisms are a criterion for the arrival of a closed-loop artificial pancreas as the present vision for the final automatic management of type 1 diabetes [7]. Hybrid modelling, data-driven techniques, and physical techniques are utilized for forecasting blood glucose levels. The data-driven method explores the present and historical values of diabetes management-linked variables to plan prospective BG excursions [8].

Generally, there are three primary categories of time-series prediction methods: classical time-series prediction, machine learning (ML), and deep learning (DL). The research reported various studies that used classical time series and ML based regression models, NNs, and DL methodologies for BG-level predictions. Eren-Oruklu et al. [9] designed an adaptive univariate autoregressive moving average (ARMA) model with fixed model orders solely based on CGM data. Meanwhile, Turksoy et al. [10] developed a multivariate ARMA with exogenous inputs (ARMAX) model, which used

exogenous variables such as glucose and insulin while excluding meal information. Given the time-varying non-stationarity of CGM data, Yang et al. proposed an ARIMA model using an algorithm that adaptively determines model orders and simultaneously updates model parameters [4]. They found that their model outperformed the adaptive univariate and ARIMA models. As more research was done, Georga et al. [11] suggested SVR and SVR with feature selection using RF and RReliefF. Hamdi et al. [12] also proposed the DE-SVR for BG-level predictions.

Researchers found that ML-based classical time series struggle to predict BG dynamics due to their assumption of linear data, sensitivity to non-stationary CGM data, difficulty in capturing complex patterns, susceptibility to noise and outliers, and inability to adapt to changing conditions. However, even though SVR and SVR with kernel functions can handle non-linearity, the limitation of machine learning methods is how the input data is represented, which has an effect on how well the predictions work.

Motivated by the excellent modelling capabilities of artificial neural networks (ANNs) for nonlinear and non-stationary phenomena, several studies have employed ANN strategies to forecast BG levels. DL differs from traditional ANNs in the number of hidden layers, their interconnections, and their ability to abstract inputs. These intricate models with increased complexity outperformed conventional shallow NNs. These advanced DL based algorithms can autonomously discern patterns from data. While feed-forward ANNs fared better at predicting BG levels, traditional RNNs have discovery constraints due to the vanishing/exploding gradient problem, which is a major drawback of using these networks.

The incorporation of memory cells and forget gates into traditional RNNs has rectified this issue in Long Short-Term Memory (LSTM) networks.

Recent studies examined RNNs and LSTMs for enhanced accuracy in BG level prediction. Fox et al. [13] employed a RNN with Gated Recurrent Unit (GRU) cells, whereas Sun et al. [14] implemented a Bi-LSTM based RNN for BG prediction. Nevertheless, the precision attained by cutting-edge models for actual patient data is insufficient, rendering these methods potentially inapplicable for diabetes care. Conventional LSTMs sometimes forfeit essential information throughout extended sequences due to the restricted capacity of their memory cells to preserve data from distant time steps. The discovery of the attention mechanism allowed researchers to address long sequencing issues and concentrate on the most pertinent segments of the sequence. Traditional LSTM models may dilute sequential information over long time spans, potentially leading the model to overlook subtle yet influential glucose patterns. The attention method enables our model to concentrate on significant previous CGM readings that may more substantially impact future glucose levels.

This study offers an Optimal Attention-Based Long Short-Term Memory for BG Level Prediction (OALSTM-BGLP) Model. At first, it employs Min-Max Scaling to normalize the input data, certifying consistent and meaningful contrasts across dissimilar features. Furthermore, the OALSTM-BGLP approach utilizes the ALSTM technique, which includes an attention mechanism to selectively concentrate on related data within the input sequence while taking into account long-term dependencies. Besides, the method has been optimized by employing the RMSProp optimizer, which adjusts the rate of learning based on the magnitude of recent gradients, facilitating efficient training and convergence. The suggested methodology outperforms state-of-the-art methods in a performance evaluation conducted on the OhioT1DM dataset.

2. LITERATURE REVIEW

For the purpose of studying and detecting diabetes mellitus, Zhang et al. [15] proposed AHDHS-Stacking, an EL framework. The two components—a feature selector (FS) and an optimiser of basic learner groupings—utilize the harmony search (HS) approach. An adaptive hyperparameter technique was used to speed up the iterative process, and the average performance of each base learner was used as an aim of FS to create the full performance. Yang et al. [16] introduced a method for abnormal blood glucose detection using an autonomous neighborhood parameter selection strategy, which is based on Adaptive Density-Based Spatial Clustering of Applications with Noise. To make up for missing blood glucose values, Yang et al. [16] proposed a feature-engineered imputation technique. Lastly, global and local spatial-temporal properties are extracted from sequence data using a temporal multi-head attention model. In the study by Butt et al. [17], a multi-layer LSTM-based RNN is used to predict BG levels in people with type 1 diabetes. The BG level is being predicted within a forecast limit using the created model.

Langarica et al. [18] suggested the Input and State Recurrent Kalman Network (ISRKN) as a new deep learning-based method for predicting glucose levels based on probabilities . In the latent space of a DNN, this network combines an input and a state Kalman filter. This lets secure calculations of later distributions and the distribution of uncertainty through Kalman calculations happen. Song et al. [19] studied a new combined ML technique termed Bagging-ABC-ELM. The artificial bee colony (ABC) system achieved the optimum input biases and weights of the ELM method. The bagging model was utilized to improve the constancy of the system and achieve greater performance than ELM. Ye et al. [20] suggested using an electric nose (E-Nose) device equipped with a metal oxide (MOX) gas sensor array as a new method to quantitatively recognize and study BG levels by determining breath biomarkers. Progressive ML algorithms have investigated and suggested exact predictions of the BG level based on the dimensions of 41 members over 10 days. Zhu et al. [21] designed an IoMT-enabled wearable device to implement the embedded model, a new DL based model that employs an attention-based evidential RNN. The device consists of a low-power, low-cost system on a chip that uses Bluetooth connectivity and edge computing to detect predictive hypoglycemia and make real-time BG predictions. Also created were desktop and cloud platforms for data storage and model fine-tuning, as well as an app for smartphones to display BG trajectories and forecasts. Lu and Song [22] suggested a new DL based hybrid technique to forecast levels of BG for 30 min of prediction horizon, that combined stacked Bi-GRU-based RNN, multi-layer perceptron, and Arora et al. [23] designed convolutional recurrent connection for 60 min aggregated CGM data for 60min prediction horizon and achieved 30.51 RMSE.

3. THE PROPOSED MODEL

In this study, we have presented a new OALSTM-BGLP methodology. The foremost intention of the OALSTM-BGLP

technique comprises different kinds of sub-processes involved in min-max normalization, ALSTM-based prediction process, and RMSProp optimizer-based parameter tuning. Figure 1 illustrates the workflow of the OALSTM-BGLP methodology.
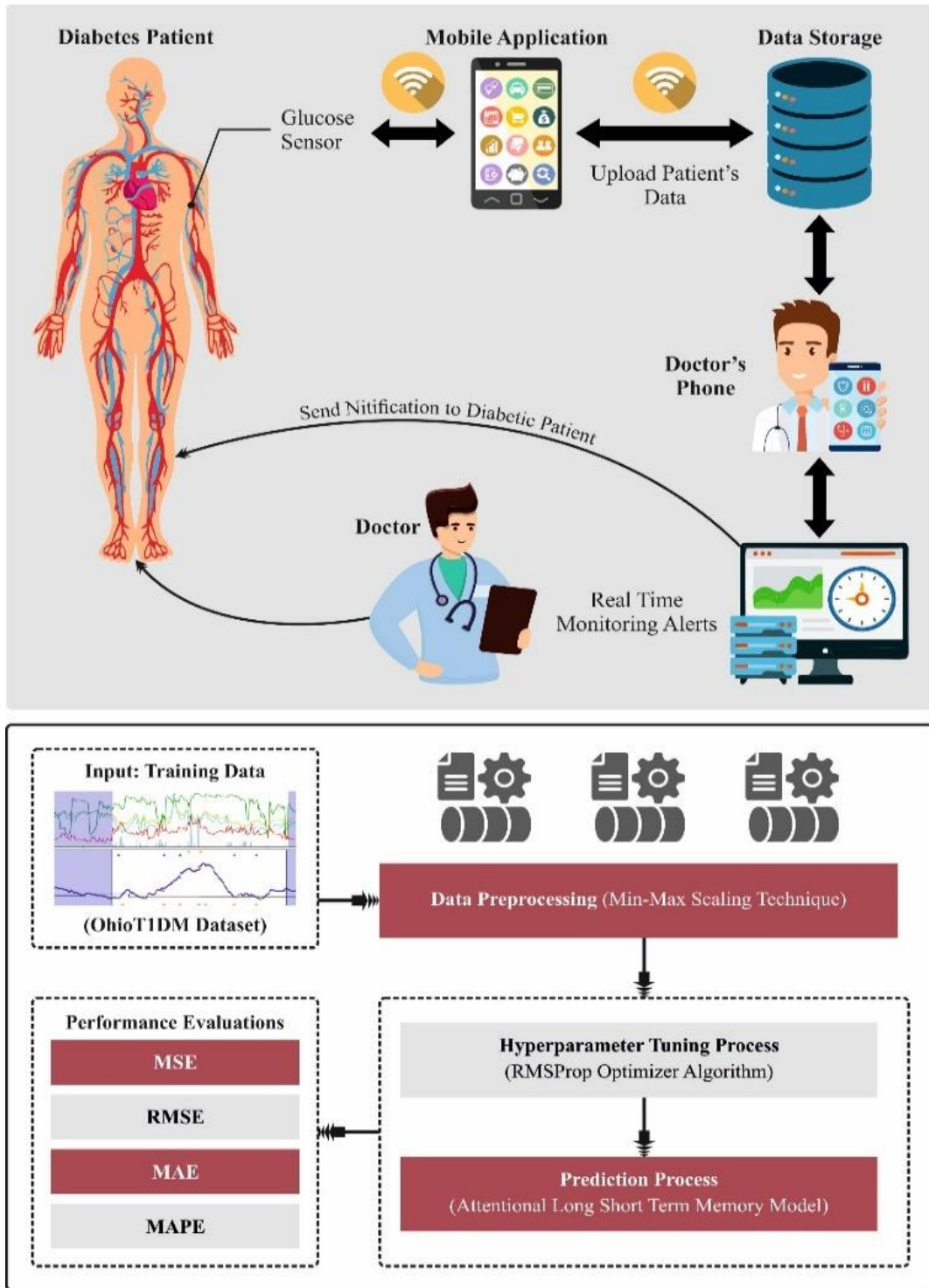

Figure 1. Workflow of the proposed OALSTM-BGLP methodology

3.1 Min-max normalization

Firstly, the OALSTM-BGLP method employs Min-Max scaling to normalize the input. Min-Max normalization is a popular data conversion method to retain the sensitive features in the data [24]. By doing so, 8 weeks real time CGM data for six patients from OhioT1DM [25] is normalized by the use of min-max normalization based on consideration of maximum and minimum values in the data. This method carries out a linear conversion on the original data. It is especially suitable

for the classification task and is used in several applications namely neural networks, artificial intelligence, clustering, nearest neighbor classification, and so on.

The goal is to normalize the original data D into a preserved form of data D' that fulfills the privacy requirement with a least possible loss with high privacy information. This technique focuses mainly on data conversion through the min-max normalization to change the original data. All the attributes in a data are normalized by mounting the value such that they fall in the minimum range within [0.0-1.0]. V_i of

attribute A from the range of [min$_i$-max$_i$] to [newmin$_A$-newmax$_A$].

$$V_i' = \frac{V_i min_A}{max_A - min_A}(\text{new}max_A\text{-new}min_A)+\text{new}min_A \qquad (1)$$

where, V_i indicates the new computed value. The relationship among the original values is preserved through the min-max normalization method.

3.2 ALSTM-based prediction process

For the prediction process, the OALSTM-BGLP technique uses the ALSTM model. In 1997, the LSTM network was initially developed [26]. Owing to its exclusive design framework, LSTM is appropriate for handling and forecasting significant actions with actual long intervals in time series. LSTM is nothing but a gated recursive neural network (GRNN) that inserts a gating device to switch the data spread in the NN based on RNN. For normal NN, the motive why performing gradient explosion is clarified below: there is a non-linear relation among the exterior condition at the latter stage and the exterior condition at this stage, and this link is assigned in every time step. The GRNN was able to resolve the problem of gradient explosion by raising the linear requirement among them. Generally, in LSTM, 3 control gate units contain namely output gate o_t forget gate f_t and input gate i_t. The gates of forget and input are the bases for LSTM to recall the dependency of long-term.

Input gate (i_t) defines how much data about the existing state wants to be kept in the interior state:

$$i_t = \sigma(U_i h_{t-1} + W_i x_t + b_i) \qquad (2)$$

where, σ denote the function of logistics, W_i and U_i represents the weight matrix; b_i refers to the bias term. h_{t-1} refers to the output of memory block at time t-1 and x_t signify the input vector at time t.

Forget gate (f_t) defines how much data from the previous wants to be rejected.

$$f_t = \sigma(U_f h_{t-1} + W_f x_t + b_f) \qquad (3)$$

where, W_f and U_f indicate the weight matrix; b_f signifies the bias.

Output gate (o_t) fixes how much data from the interior state wants to output to the exterior state at an existing stage.

$$o_t = \sigma(U_o h_{t-1} + W_o x_t + b_o) \qquad (4)$$

where, W_o and U_o denote the weight matrix; b_o refers to the bias term.

The foremost steps of the LSTM prediction method are mentioned below.

(1) At first, the exterior state at the previous stage (h_{t-1}) and the input vector at the existing stage (x_t) have been employed to compute candidate state ($\tilde{C_t}$) and 3 gates;

(2) Next, merge the input gate (i_t) and forget gate (f_t) to upgrade the interior state (C_t) of the existing stage;

(3) Finally, dependent upon the output gate (o_t) the interior state data was transmitted to the exterior state (h_t).

The abovementioned steps can recognize the prediction of the LSTM technique recognized in Keras utilizing TensorFlow.

LSTM is capable of efficiently resolving the issue of

gradient explosion [27]. At the procedure of training, overfitting will basis the test set to collapse. To resolve this issue, a dropout layer was inserted to enhance the framework of LSTM. Generally, the LSTM places the NN input vector to fixed measurement, which has reliable results while dealing with lower sizes. If the size of the input parameter is comparatively big, it will disturb the solution of the method owing to the size explosion issue. To concentrate on the prominent parameter, this research presents an attention mechanism (AM), which is recognized by preserving the LSTM intermediate output encoding for the input series.

Most attention techniques trust an Encoding-Decoding structure [28]. The encoding procedure of the original codec method produces a vector of intermediate in the technique of Seq2Seq that is utilized to keep the data of the original series. However, the vector is fixed. If the input original series length is extensive, then this vector can able to store restricted data which bounds the accepting capability of the system. Use an AM to pause the restraints of the original codec technique on a fixed vector. Figure 2 represents the infrastructure of ALSTM.
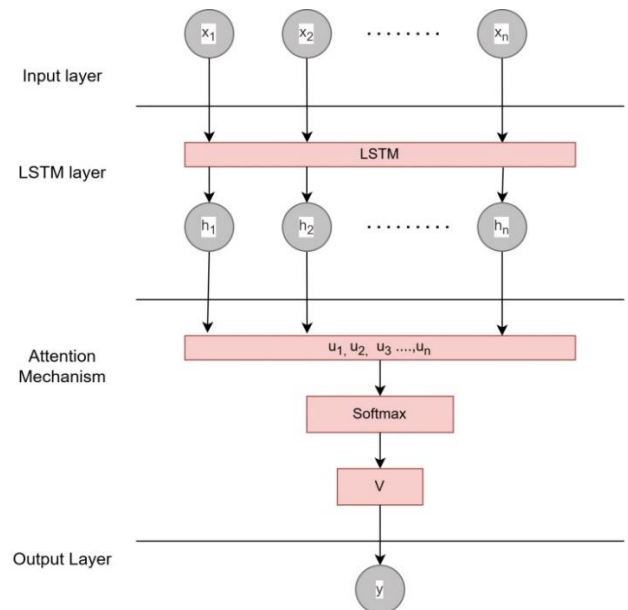

Figure 2. Structure of ALSTM

The AM is primarily executed in the subsequent steps [26]. In LSTM, the output [h_1, h_2, ..., h_n] is changed non-linearly to get [u_1, u_2, u_3, ... u_n]. When predicting BG levels using CGM data with an Attention LSTM, more attention is typically given to the most relevant past glucose readings that strongly influence the current prediction. The AM dynamically allocates higher weights to time steps in the input sequence where patterns or trends are most critical for making accurate predictions.

The significance of each intermediate time step may be represented by the attention weight matrix α_1, α_2, α_3, ..., α_n, which is generated by the AM. The weight matrix (W_k) can signify the prominence of the intermediate state. Lastly, a weighted sum has been executed for a parameter of input and weight to get the encoded vector V. The output y is acquired by decode as per the encode vector V. The complete formulation is mentioned below:

$$u_k = \tanh(W_k h_k + b_k) \qquad (5)$$

$$\alpha_k = \frac{\exp{(u_k^T u_s)}}{\sum_{k=1}^{n} \exp(u_k^T u_s)} \tag{6}$$

$$V = \sum_{k}^{n} \alpha_k \, h_k \tag{7}$$

The weight matrix is represented by W_k, a bias term or offset is denoted as b_k, normalised attention weights are represented by α_k, and the attention matrix for the time series CGM data is randomly initialised as u_s.

3.3 Hyperparameter tuning using RMSProp optimizer

Eventually, the parameter tuning process was executed by using the RMSProp optimizer to boost the ALSTM model performance. Optimizer is a hyperparameter required for training DL networks. They are widely classified into dual for non-convex optimizer issues, such as adaptive rate of learning optimization techniques, and non-adaptive rate of learning (classical SGD). Even though there is a second-order optimizer technique for a convex problem, the considerable variation among convex and non-convex optimizer issues for convex optimizer problem has one global goal, while the non-convex optimizer issue has more than one local goal. The objective is on first-order optimizer algorithms for non-convex problems. This is due to a non-convex optimization problem that is the most widespread in NN study. The loss surface defines the difficulty and complexity of finding the global optima. Therefore, an optimizer is employed for reducing the cost function which is evaluated by the cross-entropy, whereas the loss function calculates the error which shows the model efficiency. RMSprop is a gradient-based optimization method used in NN model training. A gradient of complex functions such as NN models tends to explode or vanish as the data transmits via the function. RMSprop was designed as a stochastic method for learning the mini-batch model.

RMSprop addresses the abovementioned problem through the moving average of the squared gradient for normalizing the gradients. This balances the step size (momentum), increases the step for a smaller gradient to prevent vanishing, and decreases the step for a larger gradient to prevent exploding. Geoffrey Hinton developed RMSProp, which is like an algorithm of gradient descent with motion [29]. RMSProp attempts to solve AdaGrad's entirely reducing rate of learning by utilizing a squared gradient moving average that uses the scale of current gradient descent for standardization. So, with the rate of learning upsurge, this method utilized might travel in a straight way with faster converge of higher steps.

$$E[g^2]_t = 0.9E[g^2]_{t+1} + 0.1g_t^2 \tag{8}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{(1-\gamma)g_{t-1}^2 + \gamma g_t + e}} \cdot g_t \tag{9}$$

where, g_t represents the moving average of squared gradients. γ denotes the term of decay which has a range from zero to one. In this work, the RMSProp optimizer is used to define the hyperparameter concerned in the ALSTM method. Below, we define the objective function that is used to measure the MSE.

$$MSE = \frac{1}{T}\sum_{j=1}^{L}\sum_{i=1}^{M}\left(y_j^i - d_j^i\right)^2 \tag{10}$$

Here, M and L represent the resultant values of layer and data respectively, whereas y_j^i and d_j^i denote the attained and desired sizes for j^{th} unit from the resultant layer of the system at time t respectively.

4. RESULTS ANALYSIS

The literature defines and uses a number of statistical criteria to quantify the performance of BGL predictions. In this work, we used the Mean Square Error (MSE), Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE). These measurements were prevalent in the literature, particularly for predictive tasks and specifically for forecasting BG levels.

The RMSE is a reliable measure of the prediction's accuracy, since it permits the error to be expressed in units identical to the predicted quantity. Here, the error is defined as the disparity between the predicted and observed BGL values at every time point where a BGL measurement exists within the prediction time frame [30].

Consider a sequence of BG levels (denoted y) and its expected measurements ŷ with a length equal to n. The RMSE is expressed as:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{11}$$

In most cases, the scale of the error is defined in percentage terms using the MAPE, which is given by:

$$\frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{12}$$

In this section, the performance analysis of the OALSTM-BGLP technique takes place on the OhioT1DM dataset.

The results are inspected under varying time horizons. Table 1 and Figure 3 represent the blood glucose prediction results of the OALSTM-BGLP technique under 15m prediction horizon (PH). These obtained result values exhibit that the OALSTM-BGLP technique displays effectual prediction under varying patient IDs. With patient ID of 540, the OALSTM-BGLP technique obtains MSE of 43.1915, RMSE of 6.5720, MAE of 3.5361, and MAPE of 0.0235. Next, based on patient ID of 552, the OALSTM-BGLP method gets MSE of 12.1690, RMSE of 3.4884, MAE of 2.3258, and MAPE of 0.0185. Also, with patient ID of 584, the OALSTM-BGLP algorithm gains MSE of 65.3590, RMSE of 8.0845, MAE of 3.6239, and MAPE of 0.0244. Meanwhile, based on patient ID of 596, the OALSTM-BGLP system achieves MSE of 18.5014, RMSE of 4.3013, MAE of 2.5968, and MAPE of 0.0201.

Table 2 and Figure 4 examine the BG prediction outcomes of the OALSTM-BGLP system at 30m PH. These experimental results showed that the OALSTM-BGLP method demonstrates effective prediction with varying patient IDs. According to patient ID 540, the OALSTM-BGLP system gains MSE of 34.5652, RMSE of 5.8792, MAE of 3.1300, and MAPE of 0.0207. Moreover, with patient ID of 552, the OALSTM-BGLP system obtains MSE of 12.0182, RMSE of 3.4667, MAE of 2.2911, and MAPE of 0.0188. Similarly, based on patient ID of 584, the OALSTM-BGLP technique achieves MSE of 67.5320, RMSE of 8.2178, MAE of 3.6693, and MAPE of 0.0252. Besides, with patient ID of 596, the OALSTM-BGLP algorithm offers MSE of 18.4980, RMSE of 4.3009, MAE of 2.5721, and MAPE of 0.0195, correspondingly.

Table 1. Prediction using OALSTM-BGLP under 15m PH

Patient ID	MSE	RMSE	MAE	MAPE
540	43.1915	6.5720	3.5361	0.0235
544	18.1007	4.2545	2.7271	0.0195
552	12.1690	3.4884	2.3258	0.0185
567	61.1774	7.8216	2.8162	0.0220
584	65.3590	8.0845	3.6239	0.0244
596	18.5014	4.3013	2.5968	0.0201

Table 2. Prediction outcomes of OALSTM-BGLP at 30m PH

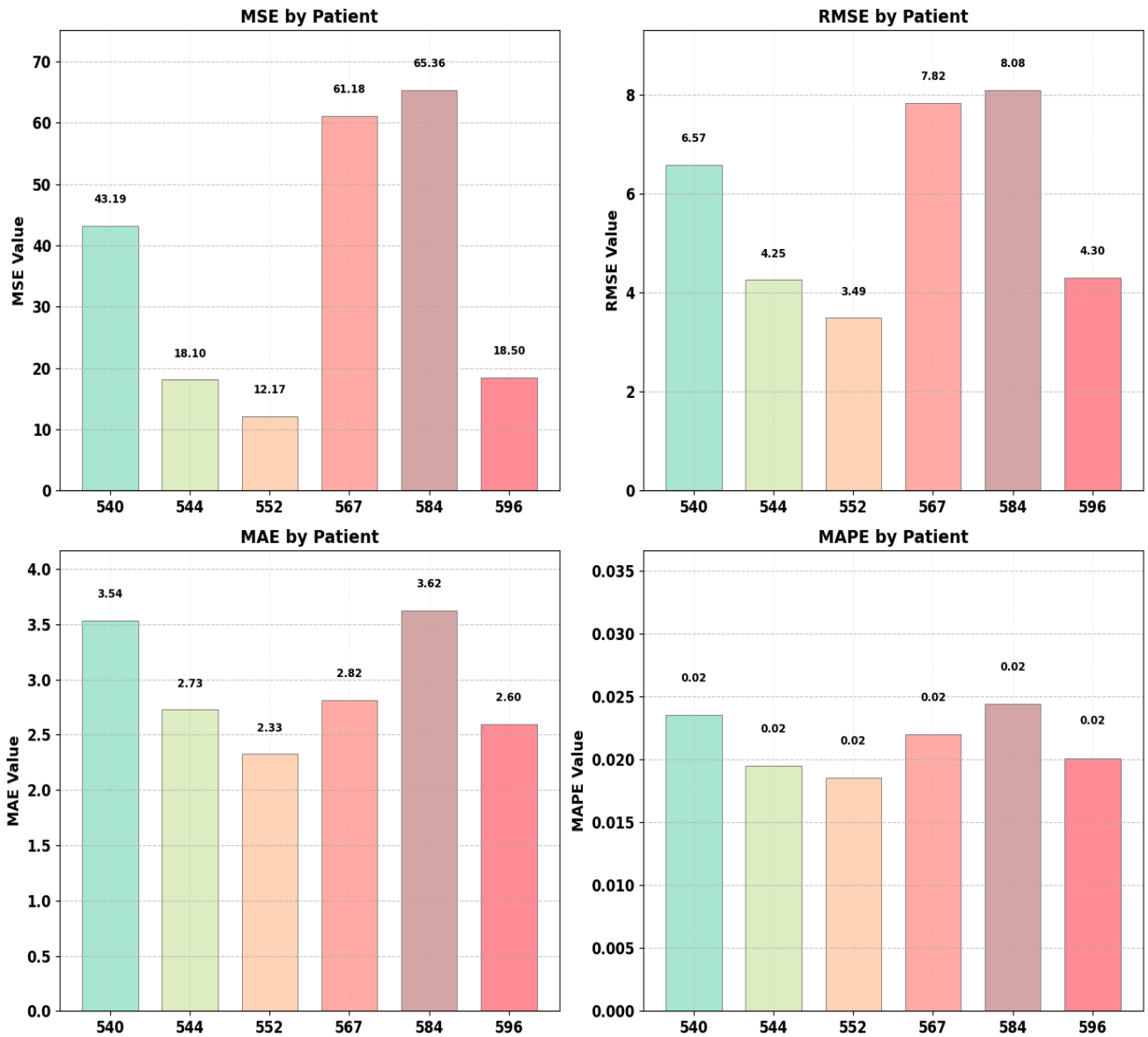| Patient ID	MSE	RMSE	MAE	MAPE
540	34.5652	5.8792	3.1300	0.0207
544	18.5069	4.3020	2.5005	0.0174
552	12.0182	3.4667	2.2911	0.0188
567	60.1172	7.7535	3.2886	0.0272
584	67.5320	8.2178	3.6693	0.0252
596	18.4980	4.3009	2.5721	0.0195

Figure 3. Prediction results of the OALSTM-BGLP model under 15m PH (a-b) MSE and RMSE (c-d) MAE and MAPE
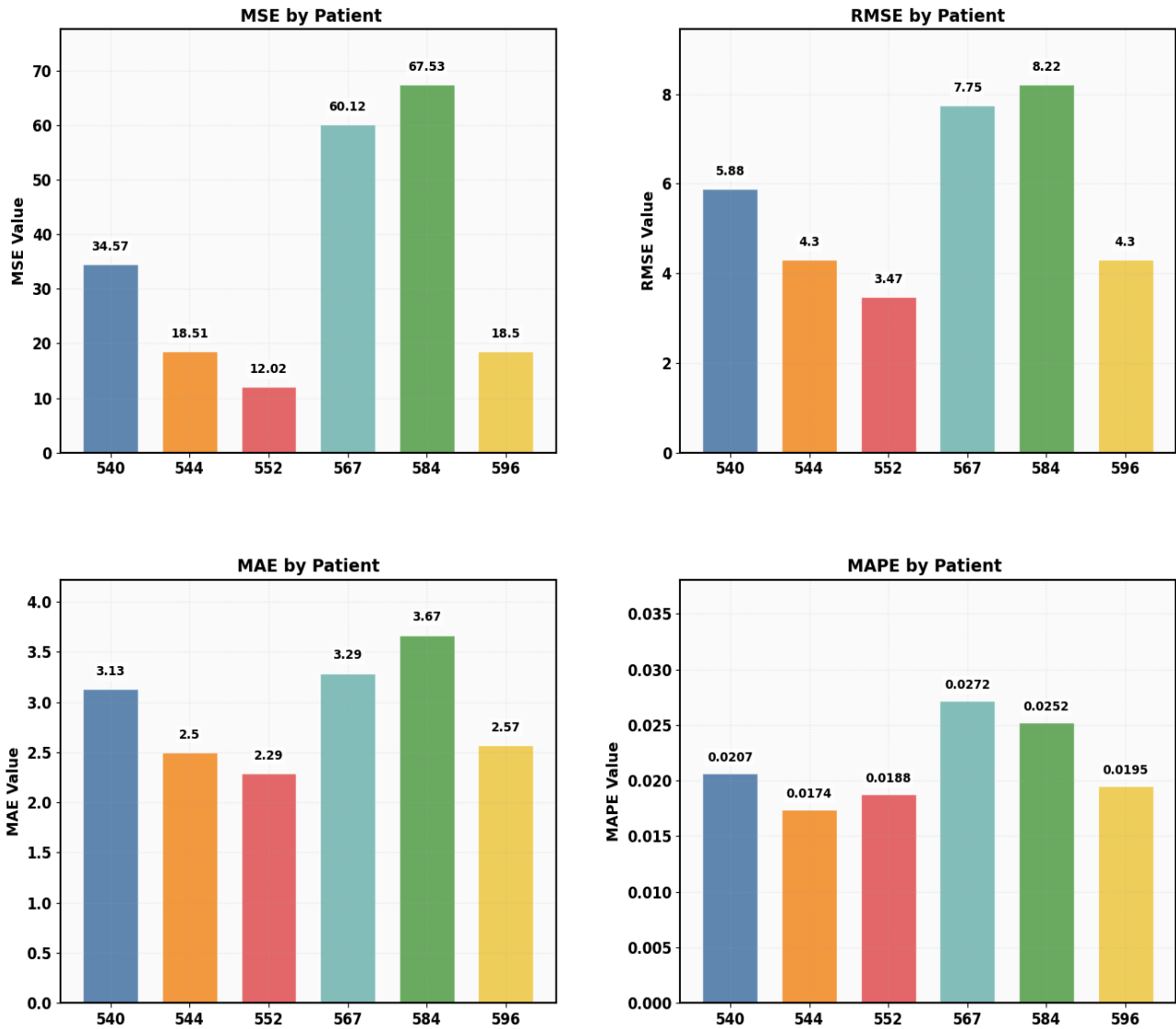
Figure 4. Prediction from the OALSTM-BGLP system at 30m PH (a-b) MSE and RMSE (c-d) MAE and MAPE

Table 3 and Figure 5 display the blood glucose prediction analysis of the OALSTM-BGLP method with 45m PH. These experimentation outcomes indicated that the OALSTM-BGLP algorithm exhibits proficient prediction at varying patient IDs. Based on patient ID 540, the OALSTM-BGLP system provides MSE of 36.2906, RMSE of 6.0242, MAE of 3.1425, and MAPE of 0.0212.

Additionally, with patient ID of 552, the OALSTM-BGLP approach gets MSE of 12.1974, RMSE of 3.4925, MAE of 2.3315, and MAPE of 0.0189. Likewise, based on patient ID of 584, the OALSTM-BGLP method acquires MSE of 68.5057, RMSE of 8.2768, MAE of 3.8021, and MAPE of 0.0260. Finally, with patient ID of 596, the OALSTM-BGLP system accomplishes MSE of 18.1017, RMSE of 4.2546, MAE of 2.5421, and MAPE of 0.0194.

Table 4 and Figure 6 demonstrate the blood glucose prediction outcomes of the OALSTM-BGLP system under 60m PH. These achieved findings denoted that the OALSTM-BGLP algorithm shows successful prediction with varying patient IDs. For the patient ID of 540, the OALSTM-BGLP method gains MSE of 39.2007, RMSE of 6.2610, MAE of 3.3479, and MAPE of 0.0227. Meanwhile, with patient ID of

552, the OALSTM-BGLP approach gets MSE of 14.0294, RMSE of 3.7456, MAE of 2.5598, and MAPE of 0.0209. Moreover, based on patient ID of 584, the OALSTM-BGLP technique provides MSE of 68.3196, RMSE of 8.2656, MAE of 3.6609, and MAPE of 0.0253. In conclusion, with patient ID of 596, the OALSTM-BGLP method achieves MSE of 20.3714, RMSE of 4.5135, MAE of 2.6548, and MAPE of 0.0202.

Table 5 and Figure 7 present a brief average result of the OALSTM-BGLP technique on varying time horizon. These experimentation findings implied that the OALSTM-BGLP technique exhibits better predictive outcomes under all horizon time intervals. With horizon time of 15m, the OALSTM-BGLP technique offers MSE of 36.4165, RMSE of 5.7537, MAE of 2.9377, and MAPE of 0.0213. Additionally, based on horizon time of 30m, the OALSTM-BGLP method gives MSE of 35.2062, RMSE of 5.6533, MAE of 2.9086, and MAPE of 0.0215. Followed by horizon time of 45m, the OALSTM-BGLP algorithm provides MSE of 36.7045, RMSE of 5.7611, MAE of 2.9926, and MAPE of 0.0222. At last, with horizon time of 60m, the OALSTM-BGLP method gains MSE of 37.5874, RMSE of 5.9081, MAE of 3.0980, and MAPE of

0.0231, respectively.

Figure 8 compares actual against predicted values of the OALSTM-BGLP system using the training dataset. The figure reported that the OALSTM-BGLP method properly predicts the blood glucose levels. It is also noticed that the predicted blood glucose levels by the OALSTM-BGLP technique are much nearer to actual values.

Figure 9 compares the actual and predicted values of the OALSTM-BGLP method on the testing dataset. This figure pointed out that the OALSTM-BGLP system appropriately predicts blood glucose levels. This can be perceived that the predicted blood glucose levels by the OALSTM-BGLP algorithm is highly close to actual values.
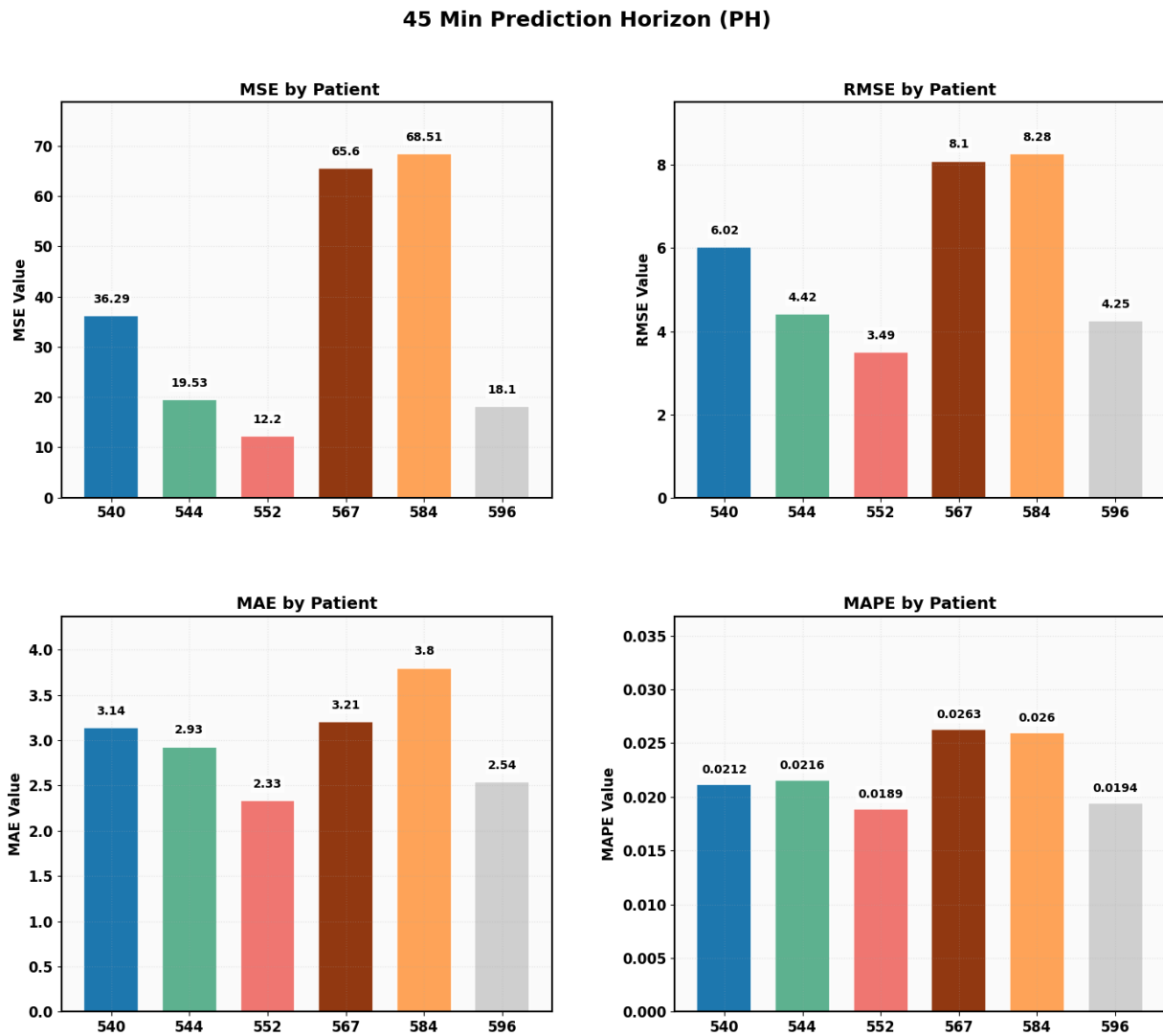
45 Min Prediction Horizon (PH)

Figure 5. Prediction of the OALSTM-BGLP model under 45m PH (a-b) MSE and RMSE (c-d) MAE and MAPE

Table 3. Prediction outcomes of the OALSTM-BGLP method on 45-min PH

Patient ID	MSE	RMSE	MAE	MAPE
540	36.2906	6.0242	3.1425	0.0212
544	19.5290	4.4192	2.9300	0.0216
552	12.1974	3.4925	2.3315	0.0189
567	65.6025	8.0995	3.2077	0.0263
584	68.5057	8.2768	3.8021	0.0260
596	18.1017	4.2546	2.5421	0.0194

Table 4. Prediction outcomes of the OALSTM-BGLP method at 60m PH

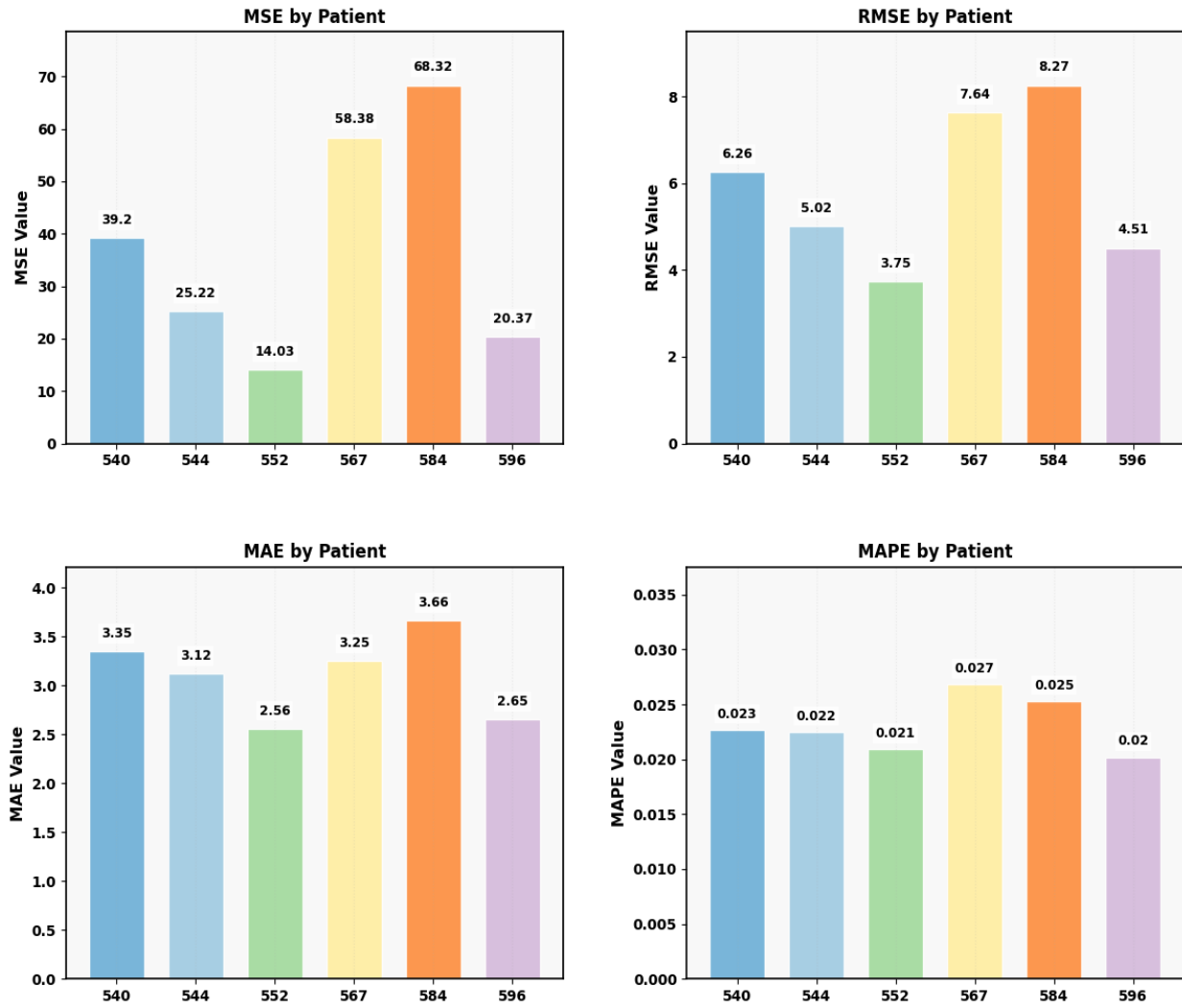| Patient ID	MSE	RMSE	MAE	MAPE
540	39.2007	6.2610	3.3479	0.0227
544	25.2215	5.0221	3.1173	0.0225
552	14.0294	3.7456	2.5598	0.0209
567	58.3816	7.6408	3.2471	0.0268
584	68.3196	8.2656	3.6609	0.0253
596	20.3714	4.5135	2.6548	0.0202

60 Min Prediction Horizon (PH)

Figure 6. Prediction outcome of the OALSTM-BGLP model under 60 PH (a-b) MSE and RMSE (c-d) MAE and MAPE

Table 5. Average PH of OALSTM-BGLP method under varying PH

PH(m)	MSE	RMSE	MAE	MAPE
15	36.4165	5.7537	2.9377	0.0213
30	35.2062	5.6533	2.9086	0.0215
45	36.7045	5.7611	2.9926	0.0222
60	37.5874	5.9081	3.0980	0.0231

Figure 10 presents the prediction concentration results of the OALSTM-BGLP technique under varying time horizons. These results pointed out that the OALSTM-BGLP technique accomplishes better performance with enhanced prediction concentration results under all-time horizons.

The results demonstrate low error values for the numerical metrics MSE, RMSE, MAE, and MAPE, indicating a significant enhancement in prediction accuracy. Precision is crucial for effective diabetes management, as accurate predictions minimise the discrepancy between predicted and actual glucose levels. This approach enhances decision-making speed by reducing errors, and aiding individuals and healthcare professionals in effectively predicting and managing hypo- or hyperglycaemic episodes. This enhances patient safety by avoiding overcorrections and facilitates personalised treatment approaches that are adapted to individual glucose variability. This increased accuracy enables

healthcare providers and patients to make informed decisions about insulin dosage, dietary changes, and lifestyle modifications. Thus, low error metrics indicate the model's robustness.

Numerical metrics are useful for evaluating model performance; however, they do not always provide a comprehensive understanding of real-world impact. Clarke's Error Grid Analysis (CEGA) is commonly utilised to evaluate the clinical reliability of blood glucose level predictions [31]. CEGA categorizes predictions into five regions: A, B, C, D, and E, concerning the clinical outcome of insulin dosing based on the predicted BG level. The most unfavorable scenario (D and E regions) involves an excessively high BG level prediction, which may result in hypoglycemia, a critical emergency condition. Consequently, we can categorise the identical absolute numerical error into different regions based on the actual BGL range and the sign of the error. A predicted

value is considered 'clinically acceptable' if it falls within either the A or B region. CEGA modifies the grid such that the 'accurate' domain matches the 'clinically acceptable' categorisation. We have generated Clarke error grids for prediction horizons at 15m, 30m, 45m, and 60m to analyze the clinical implications of the OALSTM-BGLP model. Figure 10 prominently displays the CEGA at 15m, 30m, 45m, and 60m prediction horizons, the majority of the data points also lie in region A, reinforcing the reliability and clinical relevance of our model's predictions across different time intervals. This consistent performance highlights the potential of the OALSTM-BGLP model in real-world applications for BG level prediction.
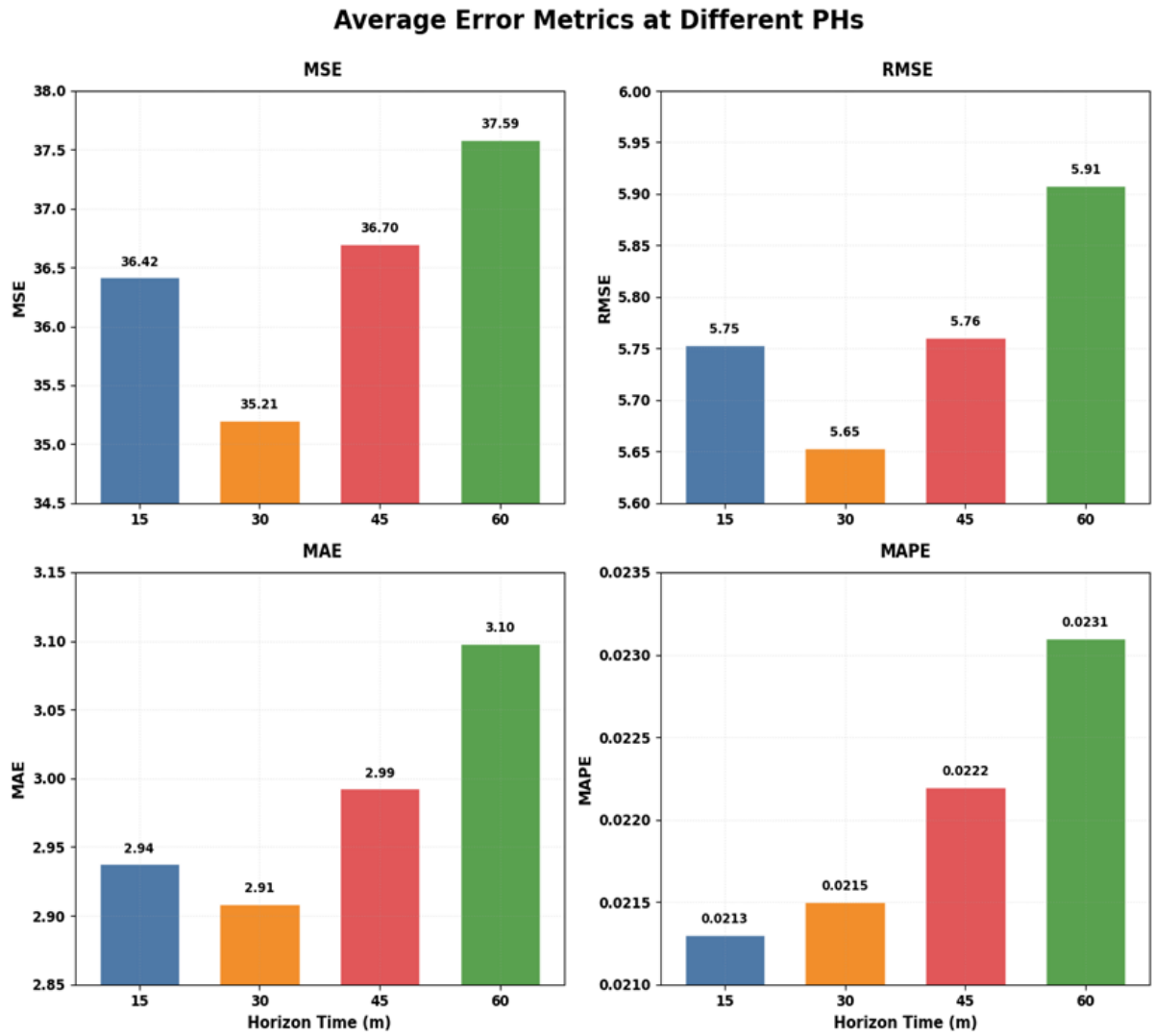

Figure 7. Average PH of OALSTM-BGLP model (a-b) MSE and RMSE (c-d) MAE and MAPE
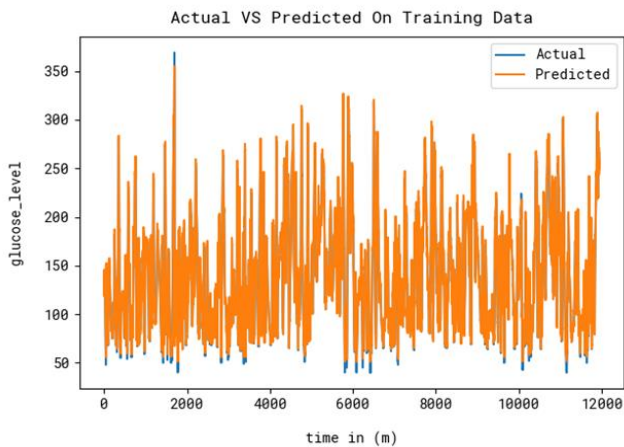

Figure 8. Comparison of actual and predicted glucose values using the OALSTM-BGLP model on the training dataset
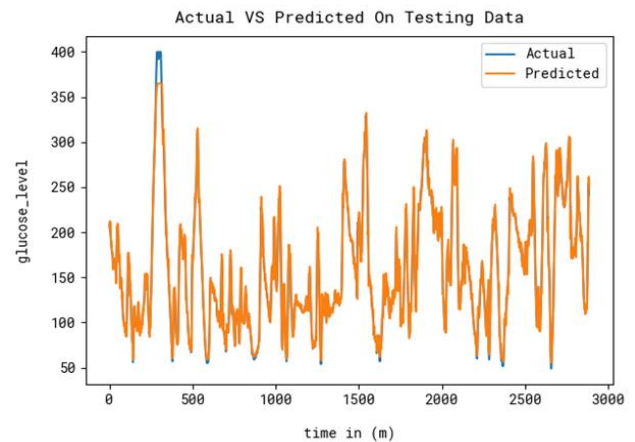
Figure 9. Comparison of actual and predicted values of the OALSTM-BGLP model under testing dataset
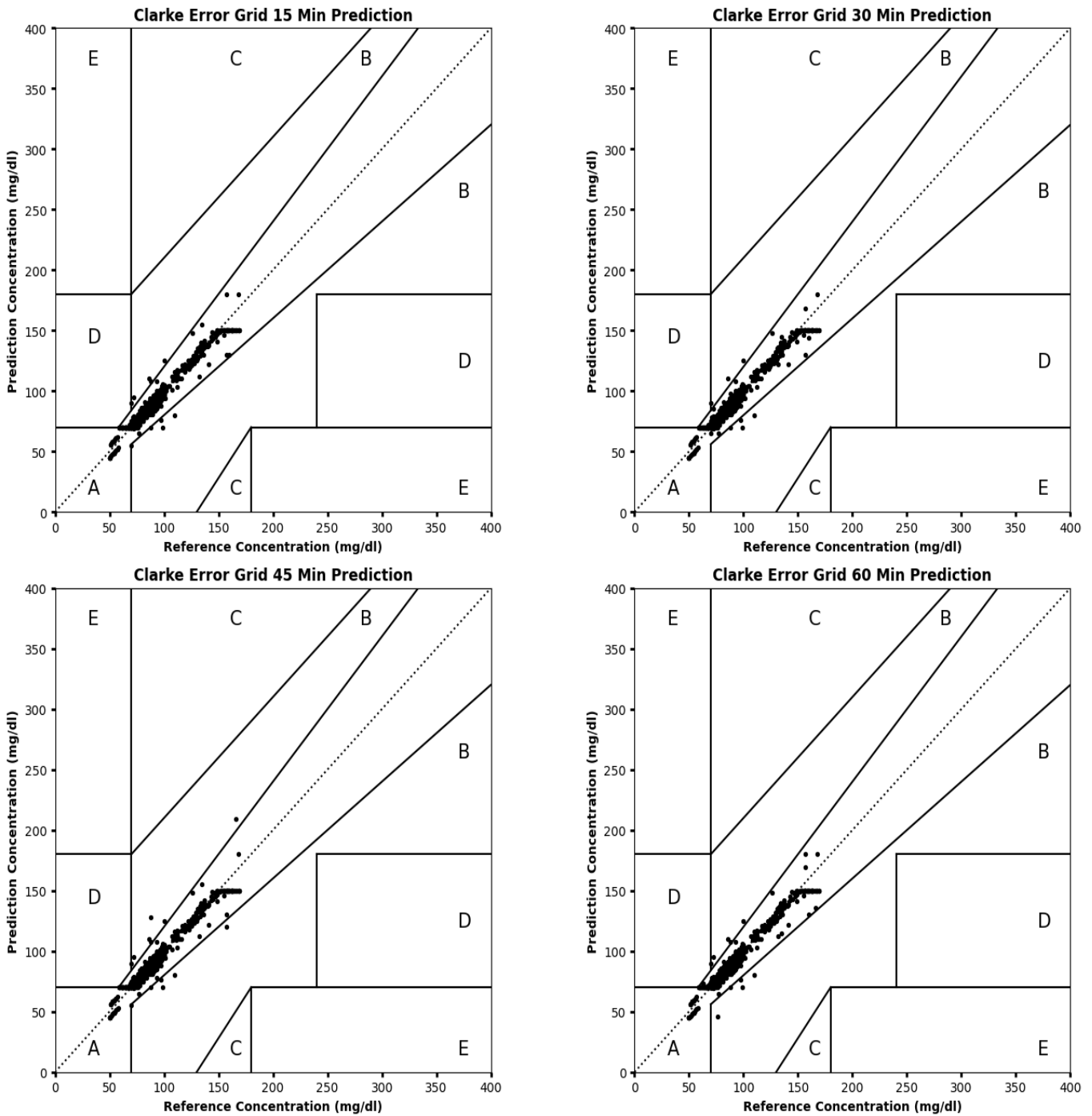
Figure 10. Clarke error grid analysis at different prediction horizons: (a) 15m, (b) 30m, (c) 45m, and (d) 60m

5. DISCUSSION

The new OALSTM-BGLP model is designed for predicting BG levels by leveraging historical CGM data across four critical prediction intervals: 15, 30, 45, and 60 minutes. The selection of time intervals for BG level prediction is determined by clinical significance and the properties of the data. Glucose levels exhibit rapid fluctuations, particularly during mealtime. Shorter intervals, such as 15 minutes, are clinically significant for monitoring rapid glucose fluctuations, facilitating prompt interventions for insulin adjustments. While intermediate intervals (30–60 m) provide valuable insights into broader trends, they are beneficial for general glucose management, including meal planning, physical activity, and insulin administration. Additionally, the sampling frequency of CGM data (every 5 m) influences the selected PH, balancing model accuracy with real-time applicability. This approach ensures the model aligns with both clinical needs and data constraints.

Our results demonstrate that the proposed model outperformed various recent DL-based approaches in terms of predictive accuracy. While direct comparison with existing studies is feasible for the 30m and 60m predictions, a noticeable gap exists in the recent literature for 15m and 45m horizons. Table 6 illustrates its performance in comparison to new models tested on various versions of the OhioT1DM dataset.

Table 6. Comparison of recent literature utilizing the OhioT1DM Dataset with the OALSTM-BGLP model

Study	Model	Patients	PHs	RMSE [mg/dL]
[32]	RNN	6	30m	22.43
			60m	36.39
[33]	Deep Residual Forecasting Architecture	6	30m	18.2
			60m	31.7
[34]	GluNet	6	30m	19.28 to 22.93 and
			60m	31.83 to 39.53
[35]	Multilayer CRNN	6	30m	17.45
			60m	33.67
[36]	DCNN, Seq-to-Seq LSTM, MLR, BRC	12	30m	17.52
			60m	24.58
[21]	E3NN	12	30m	18.92
			60m	32.54
[30]	LSTM+WaveNet+GRU	12	30m	21.90,
			45m	29.12,
			60m	35.10
[37]	Autonomous Channel DL Framework	12	30m	16.77 to 21.08 and
			60m	28.36 to 35.22
Proposed	**OALSTM-BGLP**	**6**	**15m**	**5.75**
			30m	**5.65**
			45m	**5.76**
			60m	**5.91**

Martinsson et al. [32] proposed an approach based on RNNs trained in an end-to-end fashion that required nothing but the glucose level history for the patient and predicted the BG on 30 m and 60 m PHs. The approach achieved the RMSE values of 22.43 and 36.39 for 30 and 60m PHs, respectively. Falcone et al. used data from six patients from an updated version of the OhioT1DM dataset in their study to suggest a new deep residual time series forecasting architecture. This architecture modified N-BEATS by incorporating a bidirectional LSTM network into each block in place of fully connected layers. The method included supplementary input variables, including bolus insulin and carbohydrates, and introduced three additional loss terms. The proposed deep residual forecasting method achieved average RMSEs of 18.2 for the 30 m glucose forecasting interval and 31.7 for the 60 m interval. Li et al. [34] introduced GluNet, a glucose forecasting method, utilising data from six patients in the Ohio T1DM 2018 dataset and employing deep neural networks. The findings indicated that GluNet attained RMSE values between 19.28 mg/dl and 22.93 mg/dl for 30 m PH, whereas for 60 m PH, RMSE values varied from 31.83 mg/dl to 39.53 mg/dl. Freiburghaus et al. [35] introduced a multilayer CRNN that employs multivariate data from six patients in the OhioT1DM dataset. The RMSE obtained was 17.45 for a 30 m PH and 33.67 for a 60 m PH.

Zhang et al. [36] designed four different models, including deep learning algorithms and regression models, and tested how well they could predict BG levels within 30 or 60m PH using data from 12 people with OhioT1DM. The two deep learning models are the DCNN and the Seq-to-Seq LSTM model. The two regression models are a MLR model and a BRC model. The Seq-to-Seq LSTM model exhibited superior performance in 30m ahead predictions with a 17.52 Average. RMSE, whereas the MLR model excelled in 60m-ahead predictions with a 24.58 avg. RMSE. Zhu et al. [21] proposed a GRU-based RNN model, E3NN, that incorporates an attention mechanism and evidential regression. The model was developed and evaluated on three datasets, including OhioT1DM (12 subjects). The model scored RMSE 18.92 on 30 m PH and 32.54 on 60 m PH. Dudukcu et al. [30] proposed a fusion of LSTM, WaveNet, and GRU models. Experiments were done with data from 12 patients with OhioT1DM to make

predictions on 30m, 45m, and 60m PH, as well as 21.90, 29.12, and 35.10 Avg. RMSE was achieved, respectively. Yang et al. [37] developed an autonomous channel deep learning framework. Researchers tested the framework on data from 12 patients in the clinical OhioT1DM dataset. The study's experiments revealed an RMSE range of 16.775 to 21.085 on 30 m PH and 28.36 to 35.22 on 60 m PH.

In this study, we predict BG levels at 15, 30, 45, and 60 m into the future, offering a more granular and comprehensive analysis compared to most recent studies, which primarily focus on 30 and 60 m horizons. By including shorter (15 m) and intermediate (45 m) prediction intervals, our approach captures more immediate and nuanced glucose dynamics, which are critical for timely interventions and effective diabetes management. Our RMSE results demonstrate robust performance across all horizons.

6. CONCLUSIONS

In this study, we have introduced the OALSTM-BGLP methodology, which combines Min-Max normalization, an attention-based LSTM model, and RMSProp optimizer-based parameter tuning for accurate BG level prediction. The OALSTM-BGLP technique initially normalises the input data through Min-Max scaling, facilitating consistent and meaningful comparisons among different features. The attention mechanism in the ALSTM model enables the model to concentrate on the most pertinent information within the input sequence, while also effectively capturing long-term dependencies. This enhances the model's capacity to extract significant features, thereby improving predictive accuracy. The RMSProp optimiser modifies the learning rate in accordance with the magnitude of recent gradients, thereby enhancing training efficiency and accelerating convergence. The assessment of the proposed method using the OhioT1DM dataset indicates its superior performance, surpassing recent state-of-the-art techniques regarding accuracy. The enhanced predictive capability can substantially assist in real-time diabetes management, enabling patients and healthcare providers to make timely and informed decisions. Potential

real-world applications encompass integration into personalised diabetes management systems, mobile health applications, and continuous glucose monitoring devices, where it can deliver actionable insights to avert hyper/hypoglycemic events.

Future research could look into the inclusion of additional variables, including insulin dosage, physical activity, and meal intake, to improve prediction accuracy. Furthermore, examining the model's incorporation into wearable devices for continuous, real-time monitoring could facilitate advancements in automated diabetes management. Future research should examine the model's performance across diverse patient populations to verify its robustness and applicability in different clinical contexts.

REFERENCES

[1] Antoniou, M., Mateus, C., Hollingsworth, B., Titman, A. (2024). A systematic review of methodologies used in models of the treatment of diabetes mellitus. Pharmacoeconomics, 42(1): 19-40. https://doi.org/10.1007/s40273-023-01312-4

[2] Yothapakdee, K., Charoenkhum, S., and Boonnuk, T. (2022). Improving the efficiency of machine learning models for predicting blood glucose levels and diabetes risk. Indonesian Journal of Electrical Engineering and Computer Science, 27(1): 555-562. http://doi.org/10.11591/ijeecs.v27.i1.pp555-562

[3] Li, L., Ma, Y., Jiang, J., Liu, Z., Ye, Z., Liu, S., Pu, C., Chen, C.S., Wan, Y. (2023). Machine learning models for blood glucose prediction in patients with Diabetes Mellitus: A systematic review and network meta-analysis. JMIR Medical Informatics, 11: e47833. https://doi.org/10.2139/ssrn.4401684

[4] Arora, S., Kumar, S., Kumar, P. (2023). Multivariate models of blood glucose prediction in type1 diabetes: A survey of the state-of-the-art. Current Pharmaceutical Biotechnology, 24(4): 532-552. https://doi.org/10.2174/1389201023666220603092433

[5] Ahmed, A., Aziz, S., Abd-alrazaq, A., Farooq, F., Househ, M., Sheikh, J. (2023). The effectiveness of wearable devices using artificial intelligence for blood glucose level forecasting or prediction: Systematic review. Journal of Medical Internet Research, 25: e40259. https://doi.org/10.2196/40259

[6] Saha, P., Marouf, Y., Pozzebon, H., Guergachi, A., Keshavjee, K., Noaeen, M., Shakeri, Z. (2024). Predicting time to diabetes diagnosis using random survival forests. MedRxiv. https://doi.org/10.1101/2024.02.03.24302304

[7] Khadem, H., Nemat, H., Elliott, J., Benaissa, M. (2023). Blood glucose level time series forecasting: nested deep ensemble learning lag fusion. Bioengineering, 10(4): 487. https://doi.org/10.3390/bioengineering10040487

[8] Langarica, S., Rodriguez-Fernandez, M., Núñez, F., Doyle III, F.J. (2023). A meta-learning approach to personalized blood glucose prediction in type 1 diabetes. Control Engineering Practice, 135: 105498. https://doi.org/10.1016/j.conengprac.2023.105498

[9] Eren-Oruklu, M., Cinar, A., Quinn, L. (2010). Hypoglycemia prediction with subject-specific recursive time-series models. Journal of Diabetes Science and Technology, 4(1): 25-33. http://doi.org/10.1177/193229681000400104

[10] Turksoy, K., Bayrak, E.S., Quinn, L., Littlejohn, E., Cinar, A. (2013). Multivariable adaptive closed-loop control of an artificial pancreas without meal and activity announcement. Diabetes Technology and Therapeutics, 15(5): 386-400. http://doi.org/10.1089/dia.2012.0283

[11] Georga, E.I., Protopappas, V.C., Polyzos, D., Fotiadis, D.I. (2015). Evaluation of short-term predictors of glucose concentration in type 1 diabetes combining feature ranking with regression models. Medical and Biological Engineering and Computing, 53(12): 1305-1318. http://doi.org/10.1007/s11517-015-1263-1

[12] Hamdi, T., Ben Ali, J., Di Costanzo, V., Fnaiech, F., Moreau, E., Ginoux, J.M. (2018). Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm. Biocybernetics and Biomedical Engineering, 38(2): 362-372. http://doi.org/10.1016/j.bbe.2018.02.005

[13] Fox, I., Ang, L., Jaiswal, M., Pop-Busui, R., Wiens, J. (2018). 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, pp.1387-1395. http://doi.org/10.1145/3219819.3220102

[14] Sun, Q., Jankovic, M.V., Bally, L., Mougiakakou, S.G. (2018). Predicting blood glucose with an LSTM and Bi-LSTM based deep neural network. In 2018 14th Symposium on Neural Networks and Applications (NEUREL), pp. 1-5. http://doi.org/10.1109/NEUREL.2018.8586990

[15] Zhang, Z., Lu, Y., Ye, M., Huang, W., Jin, L., Zhang, G., Ge, Y., Baghban, A., Zhang, Q., Wang, H., Zhu, W. (2024). A novel evolutionary ensemble prediction model using harmony search and stacking for diabetes diagnosis. Journal of King Saud University-Computer and Information Sciences, 36(1): 101873. https://doi.org/10.1016/j.jksuci.2023.101873

[16] Yang, G., Liu, S., Li, Y., He, L. (2023). Short-term prediction method of blood glucose based on temporal multi-head attention mechanism for diabetic patients. Biomedical Signal Processing and Control, 82: 104552. https://doi.org/10.1016/j.bspc.2022.104552

[17] Butt, H., Khosa, I., Iftikhar, M.A. (2023). Feature transformation for efficient blood glucose prediction in type 1 diabetes mellitus patients. Diagnostics, 13(3): 340. https://doi.org/10.3390/diagnostics13030340

[18] Langarica, S., Rodriguez-Fernandez, M., Doyle, F.J., Núñez, F. (2023). A probabilistic approach to blood glucose prediction in type 1 diabetes under meal uncertainties. IEEE Journal of Biomedical and Health Informatics, 27(10): 5054-5065. https://doi.org/10.1109/JBHI.2023.3309302

[19] Song, S., Wang, Q., Zou, X., Li, Z., Ma, Z., Jiang, D., Fu, Y., Liu, Q. (2023). High-precision prediction of blood glucose concentration utilizing Fourier transform Raman spectroscopy and an ensemble machine learning algorithm. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 303: 123176. https://doi.org/10.1016/j.saa.2023.123176

[20] Ye, Z., Wang, J., Hua, H., Zhou, X., Li, Q. (2022). Precise detection and quantitative prediction of blood glucose level with an electronic nose system. IEEE Sensors Journal, 22(13): 12452-12459. https://doi.org/10.1109/JSEN.2022.3178996

[21] Zhu, T., Kuang, L., Daniels, J., Herrero, P., Li, K., Georgiou, P. (2022). IoMT-enabled real-time blood glucose prediction with deep learning and edge computing. IEEE Internet of Things Journal, 10(5): 3706-3719. https://doi.org/10.1109/JIOT.2022.3143375

[22] Lu, X., Song, R. (2022). A hybrid deep learning model for blood glucose prediction. In 2022 IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS), pp. 1037-1043. https://doi.org/10.1109/DDCLS55054.2022.9858348

[23] Arora, S., Kumar, S., Kumar, P. (2025). Forecasting blood glucose level using convolutional recurrent connection. Sādhanā, 50: 16. https://doi.org/10.1007/s12046-024-02657-y

[24] Raju, K.S., Govardhan, A., Rani, B.P., Sridevi, R., Murty, M.R. (2020). Proceedings of the Third International Conference on Computational Intelligence and Informatics: ICCII 2018. Springer Nature.

[25] Marling, C., Bunescu, R. (2020). The OhioT1DM dataset for blood glucose level prediction: Update 2020. CEUR Workshop Proceedings, 2675: 71-74.

[26] Kang, Q., Chen, E.J., Li, Z.C., Luo, H.B., Liu, Y. (2023). Attention-based LSTM predictive model for the attitude and position of shield machine in tunneling. Underground Space, 13: 335-350. https://doi.org/10.1016/j.undsp.2023.05.006

[27] Pathak, D., Kashyap, R. (2023). Neural correlate-based e-learning validation and classification using convolutional and long short-term memory networks. Traitement du Signal, 40(4): 1457-1467. https://doi.org/10.18280/ts.400414

[28] Li, C., Cai, Y., Li, Y., Zhang, P. (2024). Fusion of dual sensor features for fall risk assessment with improved attention mechanism. Traitement du Signal, 41(1): 73-83. https://doi.org/10.18280/ts.410106

[29] Yaqub, M., Feng, J., Zia, M.S., Arshid, K., Jia, K., Rehman, Z.U., Mehmood, A. (2020). State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images. Brain Sciences, 10(7): 427. https://doi.org/10.3390/brainsci10070427

[30] Dudukcu, H.V., Taskiran, M., Yildirim, T. (2021). Blood glucose prediction with deep neural networks using weighted decision level fusion. Biocybernetics and Biomedical Engineering, 41(3): 1208-1223. https://doi.org/10.1016/j.bbe.2021.08.007

[31] Ali, J.B., Fnaiech, F., Saad, N., Najeh, I., Fnaiech, N., Burrus, N. (2018). Continuous blood glucose level prediction of type 1 diabetes based on artificial neural network. Biocybernetics and Biomedical Engineering, 38(4): 828-840. https://doi.org/10.1016/j.bbe.2018.08.002

[32] Martinsson, J., Schliep, A., Eliasson, B., Mogren, O. (2020). Blood glucose prediction with variance estimation using recurrent neural networks. Journal of Healthcare Informatics Research, 4: 1-18. https://doi.org/10.1016/j.bbe.2018.06.005

[33] Rubin-Falcone, H., Fox, I., Wiens, J. (2020). Deep residual time-series forecasting: Application to blood glucose prediction. KDH@ECAI, 20: 105-109.

[34] Li, K., Liu, C., Zhu, T., Herrero, P., Georgiou, P. (2019). GluNet: A deep learning framework for accurate glucose forecasting. IEEE Journal of Biomedical and Health Informatics, 24(2): 414-423. https://doi.org/10.1109/JBHI.2019.2929398

[35] Freiburghaus, J., Rizzotti, A., Albertetti, F. (2020). A deep learning approach for blood glucose prediction of type 1 diabetes. In Proceedings of the 5th International Workshop on Knowledge Discovery in Healthcare Data co-located with the 24th European Conference on Artificial Intelligence (ECAI 2020), pp. 131-135. https://ceur-ws.org/Vol-2675/paper23.pdf.

[36] Zhang, M., Flores, K.B.,Tran, H.T. (2021). Deep learning and regression approaches to forecasting blood glucose levels for type 1 diabetes. Biomedical Signal Processing and Control, 69: 102923. https://doi.org/10.1016/j.bspc.2021.102923

[37] Yang, T., Yu, X., Ma, N., Wu, R., Li, H. (2022). An autonomous channel deep learning framework for blood glucose prediction. Applied Soft Computing, 120: 108636. https://doi.org/10.1016/j.asoc.2022.108636

NOMENCLATURE

BG	Blood Glucose
BRC	Bidirectional Reservoir Computing
CGM	Continuous Glucose Monitoring
DL	Deep Learning
DCNN	Dilated Convolutional Neural Network
ML	Machine Learning
PH	Prediction Horizon
NN	Neural Networks
RNN	Recurrent Neural Network
GRNN	Gated Recurrent Neural Network
LSTM	Long Short Term Memory
AM	Attention Mechanism
ALSTM	Attention based Long Short Term Memory
SGD	Stochastic Gradient Descent
MSE	Mean Square Error
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLR	Multiple Linear Regression

Greek Symbols

α	Weight
σ	Function of Logistics
Υ	Decay Term
η	Learning Rate
θ	Parameter value w.r.t time

Subscripts

A	Attribute
t	Time