BIOCYBERNETICS AND BIOMEDICAL ENGINEERING 41 (2021) 1208–1223

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

Blood glucose prediction with deep neural networks using weighted decision level fusion

Hatice Vildan Dudukcu *, *Murat Taskiran*, *Tulay Yildirim*

Yildiz Technical University Dept. of Electronics and Communication Engineering, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 29 May 2021
Received in revised form
14 July 2021
Accepted 23 August 2021
Available online 6 September 2021

Keywords:
Blood glucose prediction
Long short term memory (LSTM)
WaveNet
Gated recurrent units (GRU)
Decision level fusion

ABSTRACT

Background and Objective: Diabetes mellitus is a chronic disease that requires regular monitoring of blood glucose in the circulatory system. If the amount of glucose in the blood is not regulated constantly, this may have vital consequences for the individual. For this reason, there are many studies in the literature that perform blood glucose (BG) prediction.

Methods: Blood glucose prediction is generally performed by using many parameters. In this paper, it was attempted to predict the future blood glucose values of the patient by using only the blood glucose values of diabetes patients' history. For this purpose, Long short term memory (LSTM), WaveNet and Gated Recurrent Units (GRU) and decision-level combinations of these architectures were used to predict blood glucose. First of all, hyper-parameters were selected for the most efficient operation of these network architectures and experimental studies were conducted using the extended OhioT1DM data set which has blood glucose history of 12 diabetes patients.

Results: Experimental studies using 30, 45 and 60 min prediction horizon (PH), the average lowest RMSE value were obtained by the fusion of three networks as 21.90 mg/dl, 29.12 mg/dl, 35.10 mg/dl respectively.

Conclusions: When the obtained RMSE value compared to state-of-art studies in the literature, the results show that the proposed method is quite successful for short-term blood glucose prediction. In addition, the proposed fusion method gives a new perspective for future studies in the literature for BG prediction.

1. Introduction

1.1. Background and objective

Diabetes mellitus is a chronic disease characterized by high glucose levels in the circulatory system. Diabetes mellitus, which has three types as type 1 diabetes, type 2 diabetes, and gestational diabetes, is a disease that can cause many health complications as well as fatal consequences. According to the International Diabetes Federation, 463 million people between the ages of 20 and 79 suffered from diabetes in 2019, and half of these people are unaware that they have dia-

* Corresponding author at: Yildiz Technical University Dept. of Electronics and Communication Engineering, 34220 Istanbul, Turkey
E-mail addresses: vdudukcu@yildiz.edu.tr (H.V. Dudukcu), mrttskrn@yildiz.edu.tr (M. Taskiran), tulay@yildiz.edu.tr (T. Yildirim).
https://doi.org/10.1016/j.bbe.2021.08.007

betes [1]. In order to cope with this disease that can cause fatal consequences, individuals must be aware of their diseases and take precautions according to the type of their diabetes. Especially, individuals suffering from Type I diabetes should measure the amount of glucose in their circulatory system at regular intervals and take certain amounts of insulin into their bodies. Type 1 diabetes occurs when an individual's pancreas cannot produce enough insulin to adjust the amount of glucose in the circulatory system. The inability of insulin to access the body at regular intervals may cause these individuals to face life-threatening situations that may result in death. For this reason, it is of great importance to know and follow the amount of glucose in the circulatory system of individuals with Type 1 diabetes. For diabetics, blood glucose monitoring is an uncomfortable procedure as it is an invasive procedure. For this reason, it is currently being studied on blood glucose monitoring with non-invasive or predictive methods instead of biochemical methods. In the study by Wolderagay et al., studies on blood glucose prediction in the literature were examined and divided into three categories as physiology-based, data-driven and hybrid approaches [2]. In this study, blood glucose prediction was performed using a deep learning fusion model as a data driven method. Detailed information about the studies conducted in the literature using machine learning models is given in the next subsection.

1.2. *Literature review*

Many studies have been conducted in the literature to predict the amount of blood glucose in the body in order to facilitate the lives of diabetes mellitus patients [3–20]. In a 2017 review study conducted by Oviedo et al. [3], the studies in the literature that have performed blood glucose (BG) level prediction for the last 5 years were examined in detail and the comparison of these studies in terms of different models, input signals and evaluation metrics was presented. The studies of blood glucose prediction that have been conducted in the literature since 2012 can be mainly examined in 3 different categories: evolutionary Methods, classical machine learning algorithms, and deep learning-based algorithms. Hidalgo et al.[4] predicted blood glucose using 10 patient data from a public hospital in Spain. In the study where different evolutionary methods were used, these methods were compared with two machine learning algorithms and it was observed that Random Forest (RF) and k-nearest neighbour (kNN) achieved regression performance up to 90%. In the study conducted by Yang et al. [11] for blood glucose prediction, Autoregressive Integrated Moving Average (ARIMA) With adaptive identification algorithms was proposed for early hypoglycemic alarms. In the experimental study using continuous glucose monitoring (CGM) data collected by the cloud data center, it was observed that the proposed method gave better results compared to the classic ARIMA model. In the study carried out by Yu et al. in 2020 [14], efficient adaptive kernel filtering algorithms were developed to reduce the computational complexity causing problems in real-time applications and to improve the blood glucose prediction in devices with hardware constraints. One of the biggest advantages of the proposed recursive kernel filtering algorithms is insensitive to abnormal CGM measurements.

The classic machine learning algorithm Support Vector Regression (SVR) and differential evolution (DE) algorithms were used together for blood glucose prediction by Hamdi et al. [8]. In experimental studies where the prediction horizon (PH) was selected as 15 min, 30 min, 45 min, and 60 min respectively, and CGM data of 12 real patients were used, the lowest average root mean square error (RMSE) was obtained as 9.44 (for 15 PH). A blood glucose prediction method based on Artificial Neural Network has been proposed by Ben Ali et al. [9]. When the results obtained in experimental studies using CGM data from 12 real patients were compared with the results of [8], It was observed that lower RMSEs were obtained for each prediction horizon. Alfian et al. [20] proposed an Extreme Gradient Boosting-based blood glucose prediction model for Type 1 diabetes in 2019. As a result of experimental studies using data from 5 patients, where continuous glucose measurement was performed for approximately 6 days, 23.2 mg/dL and 35.8 mg/dL RMSE values were obtained for prediction horizon 30 min and prediction horizon 60 min, respectively.

In recent years, especially after 2017, the most widely used method among studies that performed blood glucose prediction has been deep learning-based methods [12,7,10,13,15,16,18,19]. The modified Wavenet structure was used by Zhu et al. [7] to predict blood glucose in 2018. The modified Wavenet structure was trained for 30 min PH and experimental studies were performed using the OhioT1DM dataset [21]. As a result of the experimental studies, the average of the best RMSE values was obtained as 21.73 mg/dL. In 2019, a new method using Empirical Mode Decomposition (EMD) and Long-Short Term Memory (LSTM) was proposed by Song et al. [12]. The purpose of the proposed method is to predict blood glucose for periods ranging from 30 min PH to 120 min PH. As a result of the experimental studies using CGM data of 174 patients collected by Shanghai Sixth People's Hospital, it has been observed that the proposed method for short-term predictions yields sufficient results, but achieving high performance in long-term predictions is still a challenge. A new perspective on the problem of predicting blood glucose has been introduced by Aliberti et al. [10] In almost all of the studies in the literature, a patient's previous blood glucose data are used in training a model to predict the future blood glucose level of that patient. In the study conducted by Aliberti et al., two different models (non-linear autoregressive (NAR) neural network and long short-term memory (LSTM) networks) were used to predict blood glucose of a completely new patient by training with the blood glucose data of many different patients. While both models were successful in short-term predictions in experimental studies, it was observed that LSTM was also successful in long-term blood glucose prediction. In a 2019 study conducted by Martinsson et al. [13], they trained the model using only the patient's past blood glucose values as input to predict blood glucose, and predicted 30 min PH with 60 min PH using LSTM structure. The advantage of this study is that blood glucose prediction is performed with a single data input without computational complexity. In 2019 studies performed by Kezhi Li et al. [15,16], two different approaches for blood glucose prediction

have been proposed. In [16], blood glucose prediction was performed using 2 different datasets using a network architecture called GluNet, which consists of pre-processing, label transforms and recover, multi-layers of dilated convolutions, and post-processing components. As a result of the experimental studies performed for 30 min PH and 60 min PH, it was observed that GLuNet obtained the lowest RMSE and Mean absolute relative difference (MARD) value when compared with other state-of-art methods. In [15], a novel convolutional recurrent neural network (CRNN) has been proposed for BG prediction. This proposed method consists of 4 different components; Pre-processing, CNN, LSTM, and Signal converter. Studies performed for 30 min PH and 60 min PH have shown that the proposed method is the fastest model with the lowest RMSE architecture compared to the other methods. A proposed method combination (VMD-IPSO-LSTM), in which the variational modal decomposition (VDM) is used to decomposed the time series of patients' blood glucose concentrations and the improved particle swarm optimization (IPSO) algorithm optimizes the LSTM hyper-parameters, has been proposed by Wang et al. to improve performance in BG prediction [18]. As a result of experimental studies using data from 56 patients with diabetes mellitus and performed for 30 min PH, 45 min PH, and 60 min PH, respectively, it was observed that the proposed method combination has a lower RMSE value compared to LSTM and VDM-LSTM. In the study conducted by Zhu et al. [19], they proposed a Generative Adversarial Networks architecture, which is modified and using gated recurrent units for BG prediction. In the experimental studies performed using Ohio T1DM dataset, 18.37 mg/dL and 32.21 mg/dL RMSE values were obtained for 30 min PH and 60 min PH, respectively. In the study conducted by Xie and Wang [17] in the last period of 2020, the comparison of machine learning algorithms used in the literature for blood glucose prediction for Type 1 diabetes patients was performed. In the comparison of 11 different machine learning algorithms using the OhioT1DM dataset, a classic Autoregression with Exogenous inputs (ARX) model has the lowest RMSE value. In the study conducted by De Bois et al. in 2020 [22], T1DMS and OhioT1DM datasets were used and nine different models used in blood glucose prediction studies were compared. As a result of 30 min, 60 min, and 90 min PH predictions using 3-h patient history, the average RMSE value for 6 patients in the OhioT1DM dataset was reduced to 20.01 mg/dl.In a study conducted by Rabby et al. in 2021 [23] using the stacked LSTM method on the 2018 version of the OhioT1M dataset, Kalman smoothing technique was used on CGM data. In this study for 30 min and 60 min PH, the average RMSE values for 6 patients were obtained as 6.45 and 17.24 mg/dl respectively. In another study carried out by Sahin and Aydin in 2021 [24], 30 and 60 min PH prediction was made using the same dataset with feed-forward artificial neural networks (ANNs), and a system was developed that calculates the insulin dose needed by the patients using these predictions. For a better comparison between studies that perform BG prediction using real patient data in the literature, some specifics of these studies are summarized in the Table 1.

When the studies for BG prediction in the literature are examined, as mentioned above, deep learning-based methods seem to be the most successful methods (methods with the

Table 1 – Summary of studies using real patient data in BG prediction (BG denotes Blood glucose level, IOI denotes Injection of insulin, M denotes Meal, CI denotes Carbohydrate intake, HR denotes Heart rate, PA denotes Physical activity, TCN denotes Temporal Convolution Network.)

Ref	Subjects	Input Data	Preprocessing	Method	PH (min)	Performance Metrics
[6]	6 real	BG	N/A	LSTM	30 and 60	Negative log-likelihood (NLL), Mean squared error (MSE), Clarke error grid
[13]	6 real	BG	N/A	LSTM	30 and 60	Root mean squared error (RMSE), Surveillance Error Grid (SEG)
[15]	10 real, 10 sim	BG + IOI + CI	1D Gaussian kernel filter	CRNN	30 and 60	Root mean squared error (RMSE), Mean absolute relative difference (MARD)
[17]	6 real	BG + IOI + CI + HR	Resampling, Kalman smoothing	LSTM, TCN	30	Root mean squared error (RMSE), Temporal Gain (TG), Normalized Energy of the Second-Order ($ESOD_n$)
[24]	6 real	BG + IOI + CI + PA	Linear interpolation	ANN	30 and 60	Root mean squared error (RMSE)
[23]	6 real	BG + IOI + CI + PA	Kalman smoothing	LSTM	30 and 60	Root mean squared error (RMSE), Mean absolute error (MAE)

lowest RMSE value in BG prediction). Among the deep learning networks, it is seen that especially RNN (such as LSTM and Gated Recurrent Units (GRU)) and CNN (dilated convolution and WaveNet) are the neural networks that are often used and they give successful results for BG prediction. When these neural networks are used for the prediction problem of time series, they can perform successful regression and prediction processes due to the memory structures they contain. For this reason, in this paper using real patient data, decision level fusion of LSTM, GRU and WaveNet deep learning networks are carried out for compensation where the singular models have low prediction accuracy. Ohio T1DM dataset and its extended version [25] were used to test the performance of the proposed method. In Section 2, detailed information about the deep learning-based models used in this paper is given. In the experimental setup section, which is the third section, the dataset used in this paper, tables and figures related to the hyper-parameters and evaluation metrics and grids are presented. In Section 4, the experimental results section, table and figures related to the test results of the proposed methods and evaluation metrics and grids are given. In the conclusion part, the interpretation of the test results obtained with the proposed method and the comparison with other studies in the literature are given.

2. Methodology

Management of diabetes mellitus with the prediction of blood glucose levels is one of the most widely studied topics in the literature recently. In this paper, three different deep neural networks were used for this aim. The basic steps of the study are given in the Figs. 1 and 2. In the training process, which is the first step of the methodology, 30 min of patient BG history was used to train the model of three different deep learning networks, to predict with 30, 45 and 60 min PH. In order for model training to be carried out successfully, possible hyper-parameters for each network were tried and the parameters that obtain the best results were selected for training. After the training, models were tested with data not used in the training process. For BG prediction, 7 different prediction values were obtained for each patient by using both the prediction values of these three neural networks and the fusion of these values. In the proposed method, a decision mechanism was created by making use of the varying success of the main three neural networks according to different BG levels. During the fusion process, the networks were evaluated for each BG level separately and the fusion was performed so that the weight of the more successful ones was higher. In order to determine the BG levels used in these decisions, SEG was used. Detailed information about three different deep neural networks and the proposed fusion method used in this paper is given in the subsections.

2.1. *Deep neural networks*

In the studies examined in the previous section, it was observed that the most successful methods for BG prediction are deep neural networks. In this paper, two recurrent neural networks and one convolutional network structure were chosen to perform BG prediction. Detailed information about these network structures is given in the subsections.

2.1.1. *LSTM*

Long short-term memory (LSTM) is a type of recurrent neural network (RNN) that avoids the vanishing gradient problem. RNNs are artificial neural network structures commonly used for sequential data or time series. The main feature of RNNs is that in addition to input data, they also use system outputs of previous inputs called memory to make current predictions [26]. For every x_t input, the RNN cell takes the memory from

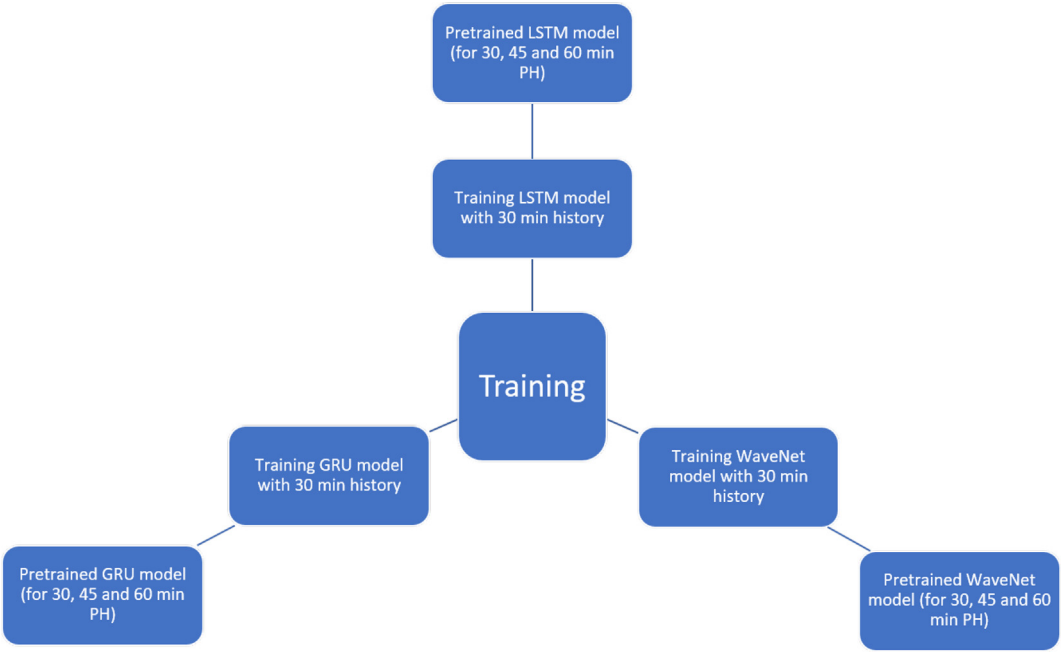


Fig. 1 – The block diagram of training process.

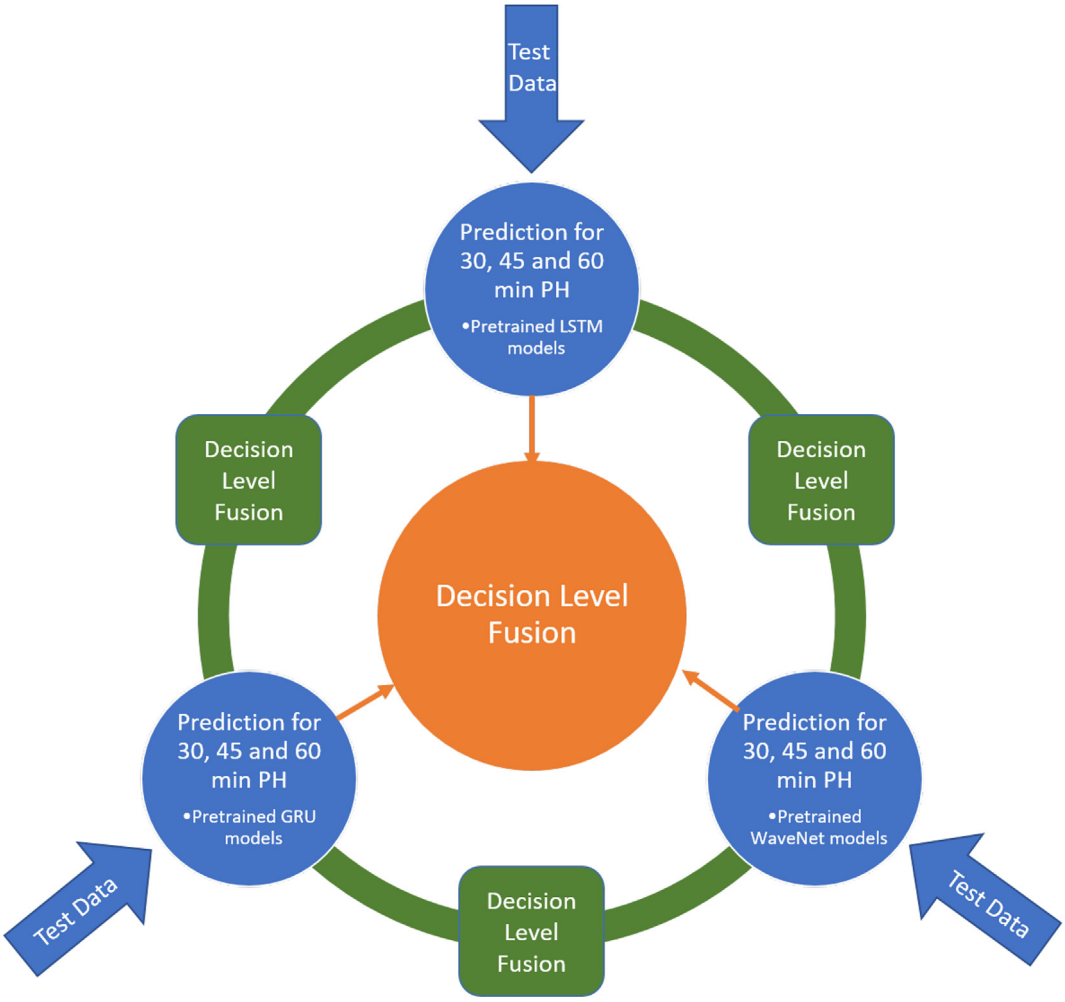


Fig. 2 – The block diagram of test process.

the previous cell h_{t-1}, and the output of the cell becomes h_t. Fort he next input x_{t+1}, the output of the previous cell h_t also becomes an input. A general RNN structure is given in Fig. 3. In traditional RNNs, the temporal evaluation of the back-propagated error is exponentially dependent on the size of the weights and this can lead to oscillating weights or learning to stop completely. This problem is called a vanishing gradient problem [27] and multiple new algorithms are proposed as a solution for this problem.

LSTM structure has gates in their cells to regulate the flow of information through their memory that processes through time [28]. An LSTM cell has three gates called input, output, and forget gates as shown in Fig. 4. The input of a cell at a given time is x_t, the previous cell memory is c_{t-1} and the output of the previous cell is h_{t-1}. The cell has two different outputs called c_t and h_t. The input and forget gates given in Eq. (1) and (2) control the cell state, which is the memory transmitted through time given in Eq. (4–5). The output gate result given in Eq. (3) gives the output of the cell also called the hidden state as seen in Eq. (6). These gates, in addition to traditional RNN cells, enable the system to overcome the vanishing gradient problem and memory to progress through time for a long time.

$$i_t = sigmoid(W_i x_t + U_i h_{t-1} + b_i) \quad (1)$$

$$f_t = sigmoid(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$o_t = sigmoid(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$\tilde{c}_t = tan(W_c x_t + U_c h_{t-1} + b_c) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_{t-1} \quad (5)$$

$$h_t = o_t \odot tan(c_t) \quad (6)$$

2.1.2. GRU

Gated recurrent unit (GRU) is a simplified and improved type of LSTM structure, with two gates called reset and update gates [29]. The input of a cell at a given time is x_t and the output of the previous cell is h_{t-1}. The calculated cell output is called h_t. Inside of a GRU cell is given in Fig. 5. Result of the hidden state is calculated by Eq. (9–10) using update Eq. (7) and reset gates Eq. (8).

$$z_t = sigmoid(W_z x_t + U_z h_{t-1} + b_z) \quad (7)$$

$$r_t = sigmoid(W_r x_t + U_r h_{t-1} + b_r) \quad (8)$$

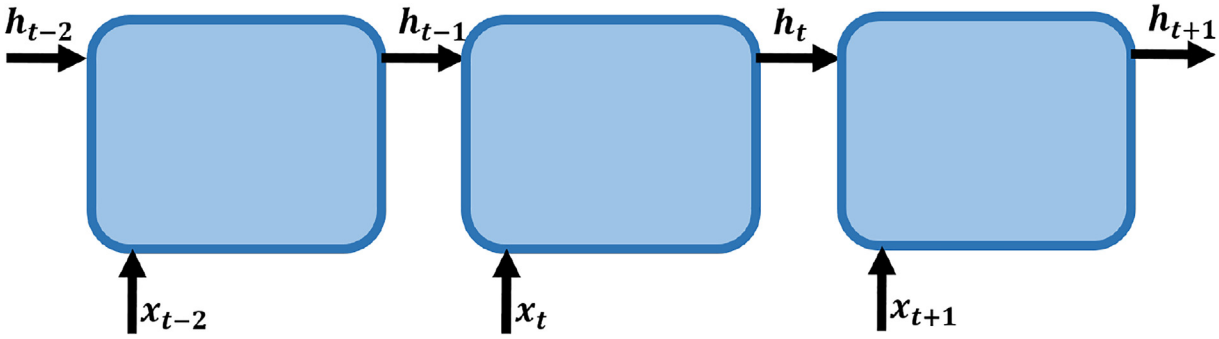


Fig. 3 – General recurrent neural network structure.

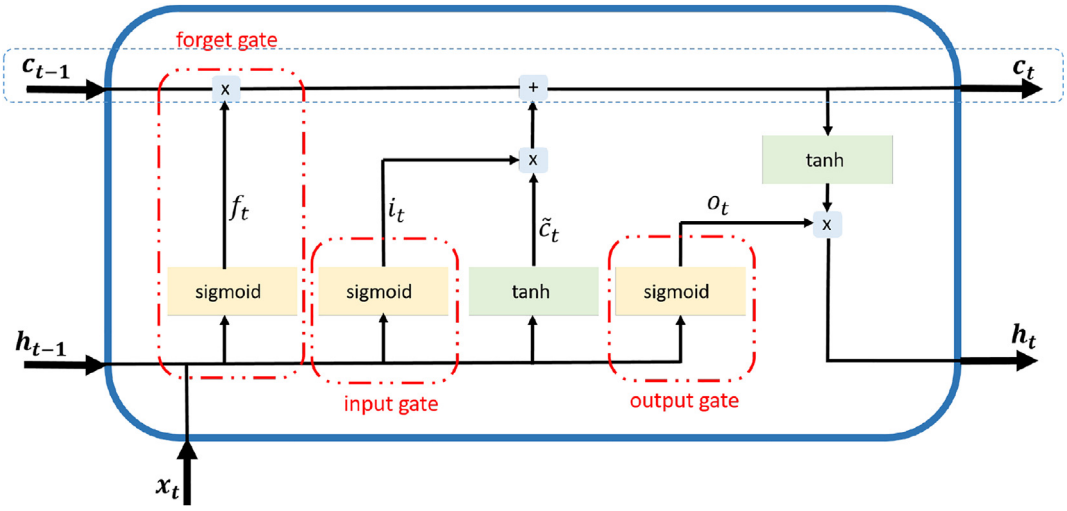


Fig. 4 – LSTM cell process structure.

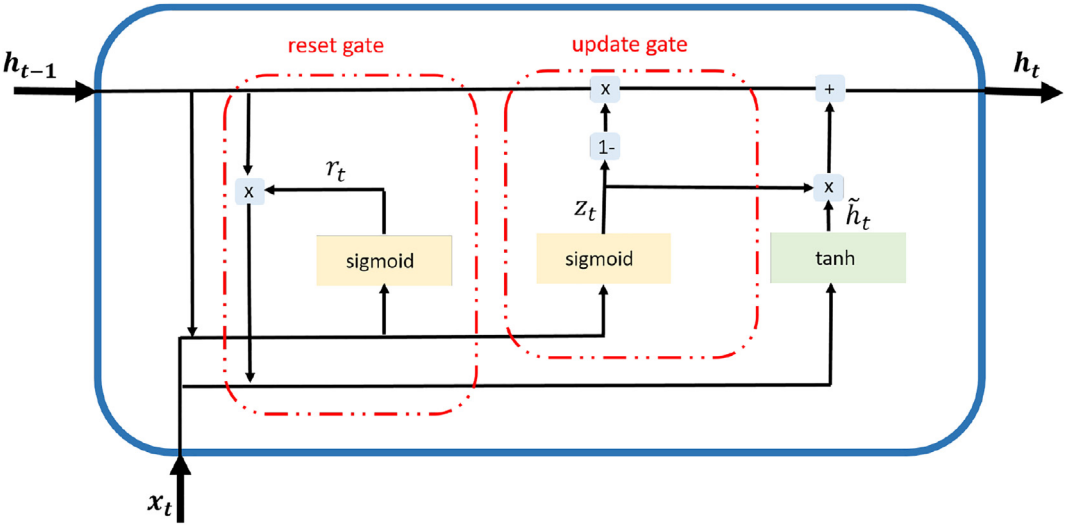


Fig. 5 – GRU cell process structure.

$$\tilde{h}_t = tan(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \quad (9)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \quad (10)$$

2.1.3. *WaveNet*

WaveNet is a one-dimensional generative convolutional neural network model first proposed for audio files [30]. Wave Net

uses a stack of convolutional layers and dilated convolution to generate the model. WaveNet network visualization is given in Fig. 6 where the dilated convolutions with dilation 1, 2, and 4 can be seen. The dilated convolution increases the receptive fields for the same input size. WaveNet uses causal convolutions and has a faster training advantage over RNNs because of the lack of recurrent connections.

2.2. *Weighted decision level fusion*

In the upper subsection, detailed information about three different deep neural networks to be used to perform BG prediction is given in this study. As seen in Fig. 1, these three neural networks were trained using 30 min patient BG history, and then BG prediction was performed for 30 min, 45 min, and 60 min PH. As seen in Fig. 2, the predictions for the three models are also passed through a rule-based decision to create decision level fusion of LSTM and GRU models, LSTM and WaveNet models, GRU and WaveNet models, and finally the fusion of LSTM, GRU and WaveNet models. Thus, seven different results were obtained for BG prediction for every patient. The purpose of using decision-level fusion method in this study is that different deep learning architectures have high performance for different glucose levels. Considering this information, with experimental studies performed using the extended version of the Ohio T1DM dataset, it was first determined which of the three different deep learning networks was successful at which blood glucose level. Surveillance Error Grid (SEG) was used to determine these blood glucose levels. Based on the information obtained, a decision level fusion mechanism has been proposed in which different deep learning architectures have lower weight in regions where blood glucose prediction fails. In the proposed fusion mechanism, the weighting values used for the effects of different deep learning networks were determined according to their performance in the high risk regions. While the fusion of two models was performed, the prediction result of that model was determined as the result of fusion in regions where one model was more successful (in more than 80% of the patients). In regions where there is no definite success of any model, the prediction result of the neural network, which is achieving higher performance than the other network (in 50%–80% of the patients) is multiplied by 0.7 coefficient, and the prediction result obtained from the other neural network is multiplied by 0.3 coefficient and the weighted sum determined as BG prediction result. While the fusion of three networks was performed, the results of these networks were determined as the result of fusion in regions where two networks show the same region. In regions where this rule could not be achieved, a weighted sum was calculated according to the accuracy of networks in BG prediction and the result of the weighted sum of three networks was determined as the result of fusion. The weight coefficients are chosen for LSTM, WaveNet, and GRU networks were determined as 0.4, 0.2, and 0.4 respectively. Fusion weight coefficients were decided by analyzing which deep learning architectures were successful in which regions after observing the performance of deep learning architectures for 30 min, 45 min, and 60 min PH. Which deep learning model was more successful was decided by looking at the RMSE value and the amount of prediction output falling into the high risk area. After this, the multiplication coefficients were tried at 0.1 intervals and the most successful results were determined as fusion coefficients. The fusion weight coefficients in this study were determined for the most comprehensive diabetes dataset in the literature in order to show the researchers that fusion is a successful method for BG prediction, and they may change if more comprehensive datasets are used in future studies. In order for the researchers to realize the system, detailed information about the fusion process is also given in Fig. 7. These fusion processes and coefficients were determined as a result of the analysis of the data of 12 patients in the OhioT1DM dataset.

Detailed information about the experimental setup created to test the BG prediction performance of the proposed method is given in the next section.

3. Experimental setup

In this paper, an experimental setup has been prepared in order to perform the BG prediction training and tests with the proposed methods given in Figs. 1 and 2. The OhioT1DM dataset [21,25] was chosen to train deep learning models and test the proposed methods. Model trainings of neural net-

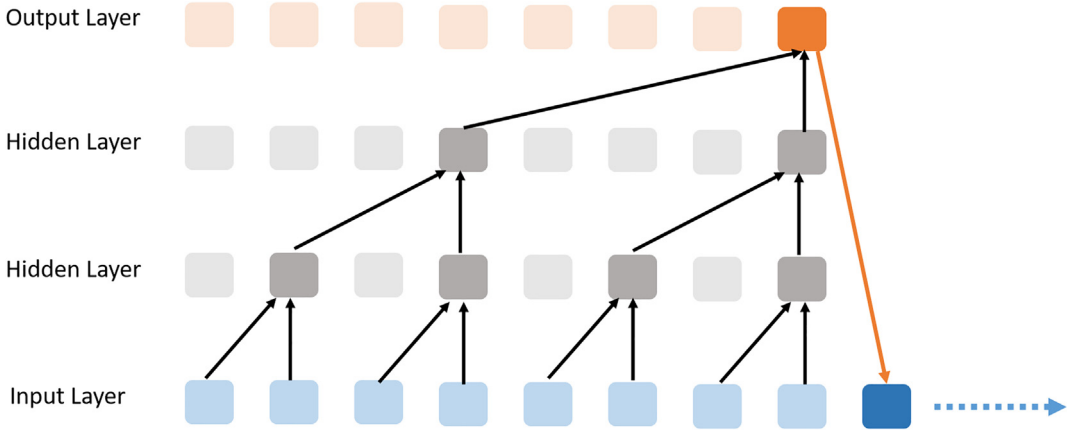


Fig. 6 – The dilated convolution in a WaveNet structure.

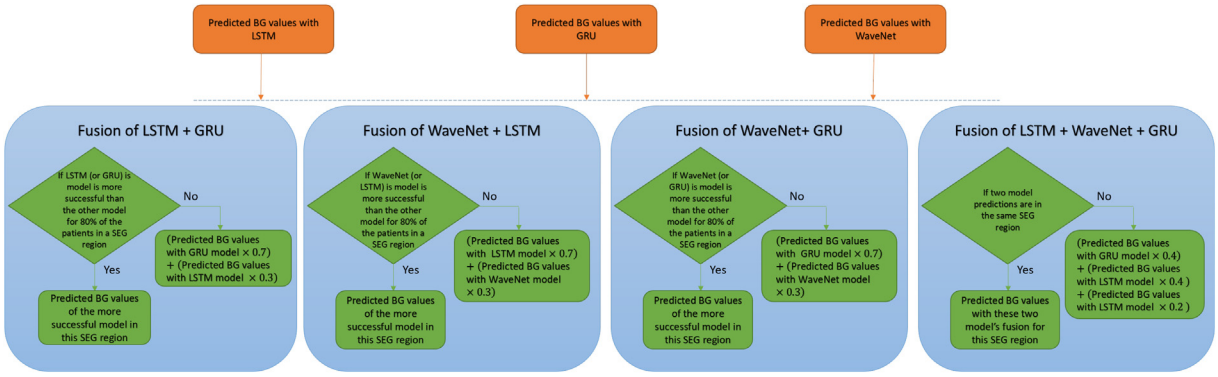


Fig. 7 – Detailed information about the performed fusion processes.

works in this study were carried out on Python software using Keras Application Programming Interface. In the same way, the tests were carried out on Python using the test datasets with the trained models. The selection of the hyper-parameters used during the training of models aims to increase the system performance. In order to evaluate the results of the experimental studies carried out within the framework of the experimental setup and for comparison with similar studies in the literature, the evaluation metrics that are widely used in recent studies have also been used in this paper. The subsections of this section contain detailed information about the dataset used in experimental studies, hyper-parameter selection of the networks used for BG prediction, and evaluation metrics used on the experimental results obtained from test data.

3.1. *Dataset*

In this paper, the OhioT1DM Dataset[21,25], which is widely used in the literature for blood glucose prediction, was chosen as the data set to be used in conducting experimental studies. This data set was first collected for the Blood Glucose Level Prediction (BGLP) Challenge in 2018 [21], and this version of the dataset contains data from 6 people suffering from type 1 diabetes. This data includes 19 different data types, especially the measurement of glucose levels in the circulatory system of the patients for 8 weeks, the amount of insulin taken by the patient, and the data obtained from physiological sensors. The first version of the dataset consists of 4 female subjects and 2 male subjects aged between 40 and 60. In 2020, the number of subjects was increased from 6 to 12 and the data set in 2018 was expanded. In the last version of the dataset[25], data were obtained from 7 male and 5 female subjects aged between 20 and 80 years. Detailed information on the OhioT1DM dataset is given in Table 2.

3.2. *Hyper-parameter selection*

Deep neural networks, which are widely preferred in solving many problems in the literature, have also been among the most preferred methods in BG prediction. One of the most important factors affecting the problem solving success of deep learning structures is the selection of hyper-parameters. In this paper, the optimum hyper-parameters of three different deep neural networks selected to carry out experimental studies were obtained and the training were carried out with them. For this purpose, grid search was used for hyper-parameter selection, where all combinations were tested by establishing a separate model for all hyper-parameter values desired to be tested, and the most successful hyper-parameter set was determined according to the RMSE values [31]. In this paper, the tried values of the hyper-parameters used for grid search were chosen from the values commonly used in the literature. As a result of

Table 2 – Comprehensive Description on The OhioT1DM Dataset [25].

ID	Gender	Age Range	Pump Model	Sensor Band	Year	Training Examples	Test Examples
540	Male	20–40	630G	Empetica	2020	11947	2884
544	Male	40–60	630G	Empetica	2020	10623	2704
552	Male	20–40	630G	Empetica	2020	9080	2352
567	Female	20–40	630G	Empetica	2020	10858	2377
584	Male	40–60	530G	Empetica	2020	12150	2653
596	Male	60–80	530G	Empetica	2020	10877	2731
559	Female	40–60	530G	Basis	2018	10796	2514
563	Male	40–60	530G	Basis	2018	12124	2570
570	Male	40–60	530G	Basis	2018	10982	2745
575	Female	40–60	530G	Basis	2018	11866	2590
588	Female	40–60	530G	Basis	2018	12640	2791
591	Female	40–60	530G	Basis	2018	10847	2760

such a hyper-parameter selection process, all combinations of the determined values were tried and the most successful hyper-parameter set was determined. Different parameters tried for LSTM network hyper-parameter selections using grid search are given in Table 3. The final version of the LSTM network to be used in training has an LSTM layer with 256 units, the activation function of the output layer is chosen as Rectified Linear Unit (ReLU) and the loss function of the network is chosen as mean squared error. The training is performed with batch size 1024 using optimizer Adam for a maximum of 10000 epochs with an early stopping criterion set to 250 epochs.

Different parameters tried for WaveNet network hyper-parameter selections using grid search are given in Table 4. The final version of the WaveNet network to be used in training has 256 and 128 units respectively, activation function of the layers is chosen as Rectified Linear Unit (ReLU) and the loss function of the network is chosen as mean squared error. The training is performed with batch size 1024 using optimizer Adam for maximum of 10000 epochs with an early stopping criterion set to 250 epochs.

The GRU network to be used in training has 2 GRU layers with 256 and 128 units respectively, activation function of the layers is chosen as Rectified Linear Unit (ReLU) function and the loss function of the network is chosen as mean squared error. The training is performed with batch size 1024 using optimizer Adam for a maximum of 10000 epochs with an early stopping criterion set to 250 epochs.

Detailed information about the metrics and grids that were used to evaluate the results of the tests performed with the OhioT1DM Dataset of the models trained with the selected hyper-parameters is given in the next subsection.

3.3. *Evaluation metrics and grids*

Evaluation metrics are standards used to measure the accuracy of systems designed to solve a problem. Many metrics and grids are used to evaluate the accuracy of BG prediction system studies conducted in the literature on BG prediction. Among these, the most commonly used numerical metric is the root mean squared error (RMSE), while the most recently used grid is Surveillance Error Grid (SEG). In this paper, four different numerical evaluation metrics, namely root mean squared error (RMSE), mean absolute error (MAE), root mean squared percentage error (RMSPE), mean absolute percentage error (MAPE), and Surveillance Error Grid (SEG) as a visualization method used to perform a comprehensive evaluation of the proposed methods [32,33]. The calculations of these numerical evaluation metrics were given in Eqs. (11)–(14). Surveillance Error Grid (SEG) is a metric for error and risk assessments of blood glucose measurements or predictions [34]. In this error grid risk ranges are identified from 0 (none) to 4 (extreme) [35]. The visualization of a simple SEG figure is given in Fig. 8. The limits of the SEG are 0 to 600 mg/dl, and the regions are separated by 120 mg/dl intervals. The ground truth data and the corresponding predicted data are then marked on the error grid. In this paper, the region distributions of the predictions are also examined. Another metric to be used in the evaluation of this study is Prediction-Error Grid Analysis (PRED-EGA), which was created by Sivananthan et al. for the evaluation of blood glucose prediction in 2011 [36]. In PRED-EGA, evaluation is performed by dividing the BG data into hypoglycemia, normoglycemia and hyperglycemia levels. Evaluation metrics and grids obtained as a result of experimental studies with the OhioT1DM Dataset are given in the experimental results section.

Table 3 – Hyper-parameter selection for LSTM network. (Hyper-parameters selected for use in training process are shown in bold.)

Units	Activation	Loss	Optimizer	Batch Size
64	Linear	**Mean Squared Error**	RMSprop	128
128	Sigmoid	Mean Absolute Error	**Adam**	256
256	**ReLU**	Mean Squared Logarithmic Error	SGD	512
512	Tanh	Mean Absolute Percentage Error	Adadelta	**1024**
128–64			Adagrad	2048
256–128			Adamax	
512–256			Nadam	

Table 4 – Hyper-parameter selection for WaveNet network. (Hyper-parameters selected for use in training process are shown in bold.)

Filters	KernelSize	MaximumDilationRate	Units	Activation	Loss	Optimizer	BatchSize
32	3	16	128–64	Linear–Linear	**MSE**	RMSprop	128
64	**4**	32	**256–128**	**ReLU-ReLU**	MAE	**Adam**	256
128	5	64	256–64	ReLU-Linear	MSLE	SGD	512
256	6	128		Linear-ReLU	MAPE	Adadelta	**1024**
512		**256**				Adagrad	2048
1024						Adamax	
						Nadam	

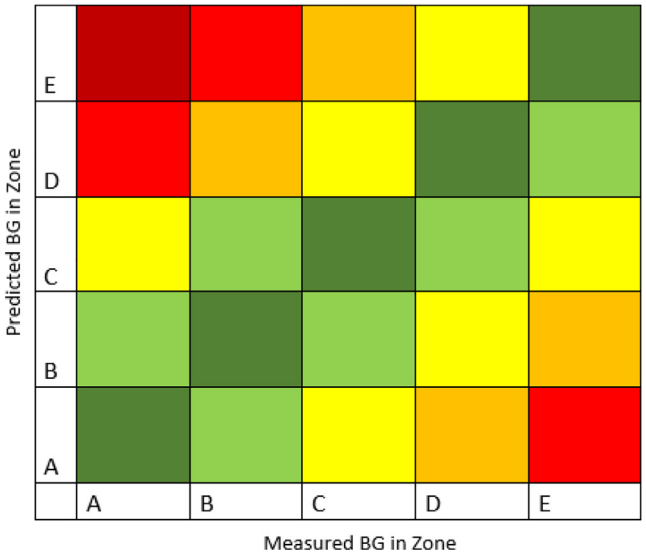


Fig. 8 – Visualization of SEG regions.

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(predicted_i - measured_i)^2} \quad (11)$$

$$MAE = \frac{1}{n}\Sigma_{i=1}^{n}|predicted_i - measured_i| \quad (12)$$

$$RMSPE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\left(\frac{predicted_i - measured_i}{measured_i}\right)^2} \quad (13)$$

$$MAPE = \frac{1}{n}\Sigma_{i=1}^{n}\frac{|predicted_i - measured_i|}{measured_i} \quad (14)$$

4. Experimental results

In this paper, LSTM, GRU, and Wavenet structures are trained with the OhioT1DM dataset, then tested with the trained models and decision model fusion as a combination of these models. The results of these tests are visualized using SEG and compared using average RMSE, MSA, MAPE, and RMSPE metrics of the patients. The dataset used in this study contains continuous blood glucose measurement data with 5 min intervals collected from 12 different patients. The predictions and the training are performed using only the measured glucose levels from the dataset. While performing the training process, 20% of the patient data for the training is split for the validation process. For every patient, training is performed using 30 min measurement history (sliding window size of 6 time-steps), and the predictions are carried out for 30 min, 45 min, 60 min PHs. While the onset of insulin occurs within 10–30 min, and the onset of meal responses on glucose levels occurs approximately within 5–10 min, knowing future BG levels at least 30 min ahead of time can prevent potentially dangerous glycemic episodes and for this prediction, the using the patient's last 30-min history are considered to contain sufficient information. In addition, with using 45 and 60 min PH it is possible to create a history from predicted values for a certain period rather than taking continuous measurements from the patient in future studies. If accurate blood glucose predictions are made with the history created with these predicted values, it can be used to determine the amount of insulin the patient needs in the future or to determine alarm conditions without the need for continuous glucose monitoring.

While the training and test process is being carried out, the missing data in The OhioT1DM dataset due to lack of

Table 5 – Average prediction performance metrics obtained from 30 min PH using the OhioT1DM dataset (Best results are shown in bold.)

Method	RMSE	MAE	RMSPE (%)	MAPE (%)
LSTM	22.13	16.02	15.46	11.01
WaveNet	22.49	16.47	15.97	11.45
GRU	22.00	15.91	15.48	10.98
WaveNet + LSTM	22.35	16.29	15.87	11.35
WaveNet + GRU	22.21	16.15	15.80	11.26
LSTM + GRU	21.98	**15.86**	**15.34**	**10.90**
LSTM + WaveNet + GRU	**21.90**	15.87	15.40	10.96

Table 6 – Average prediction performance metrics obtained from 45 min PH using the OhioT1DM dataset (Best results are shown in bold.)

Method	RMSE	MAE	RMSPE (%)	MAPE (%)
LSTM	29.28	21.61	**20.84**	14.94
WaveNet	29.68	22.19	21.68	15.60
GRU	29.22	**21.50**	20.88	**14.88**
WaveNet + LSTM	29.46	21.87	21.20	15.28
WaveNet + GRU	29.44	21.83	21.27	15.28
LSTM + GRU	29.26	21.56	20.88	14.92
LSTM + WaveNet + GRU	**29.12**	21.52	20.86	14.93

Table 7 – Average prediction performance metrics obtained from 60 min PH using the OhioT1DM dataset (Best results are shown in bold.)

Method	RMSE	MAE	RMSPE (%)	MAPE (%)
LSTM	35.17	**26.39**	**25.76**	**18.42**
WaveNet	35.50	27.00	26.69	19.20
GRU	35.31	26.51	25.97	18.54
WaveNet + LSTM	35.31	26.73	26.26	18.90
WaveNet + GRU	35.43	26.79	26.41	18.95
LSTM + GRU	35.30	26.45	25.87	18.46
LSTM + WaveNet + GRU	**35.10**	26.41	25.89	18.53

Table 8 – Numerical evaluation metrics of each patient for the LSTM + WaveNet + GRU fusion model (30 min PH).

Patient ID	RMSE	MAE	RMSPE (%)	MAPE (%)
540	25.28	18.77	17.86	13.10
544	19.76	14.36	13.42	09.61
552	19.43	14.66	15.07	11.32
559	21.78	15.35	15.19	10.23
563	20.43	14.39	12.72	08.96
567	23.96	17.41	16.63	12.75
570	18.06	12.85	09.61	06.59
575	25.02	16.77	18.15	12.37
584	24.84	18.57	18.61	12.78
588	21.26	15.55	12.83	09.38
591	23.76	18.16	20.29	14.45
596	19.23	13.58	14.44	09.95
Average	**21.90**	**15.87**	**15.40**	**10.96**

measurement data up to a few hours were extracted to ensure the continuity of the prediction during the test and train process. This process was done by creating a series with uninterrupted data for 30 min history and then the required PH time period from 5 min intervals of data, and in the case of series having missing data, this time series was not included in the training and test datasets. After checking the continuity of the time series, the series was divided into windows containing the last 6 time-steps, and these windows were given as 30 min history input to the neural network and prediction was performed for 30, 45, 60 min PH. The reason why the data is not filled artificially with some techniques present in the literature is that some data loss is approximately twenty-four hours and the artificial data produced may be meaningless for such a large period of time. The optimum hyperparameter combinations for LSTM, GRU, and WaveNet structures were used for the training process, and obtained pretrained neural networks were evaluated with Ohio T1DM dataset. For every patient, training data is first normalized then the LSTM, GRU, and WaveNet models are trained with 30 min measurement sequence. The training process was carried out for each patient model using the patients' own BG history. The trained models are evaluated with the test data for the patients. These three neural networks were trained

Table 9 – SEG average risk range distribution ratios (%) for 30 min PH using the OhioT1DM dataset (Best method is shown in bold.)

Method	None (0)	Slight (1)	Moderate (2)	Great (3)	Extreme (4)
LSTM	86.42	13.56	0.02	0.00	0.00
WaveNet	85.91	14.06	0.03	0.00	0.00
GRU	86.41	13.55	0.04	0.00	0.00
WaveNet + LSTM	86.43	13.54	0.03	0.00	0.00
WaveNet + GRU	86.39	13.57	0.04	0.00	0.00
LSTM + GRU	86.52	13.44	0.04	0.00	0.00
LSTM + WaveNet + GRU	**86.53**	**13.45**	**0.02**	**0.00**	**0.00**

Table 10 – SEG average risk range distribution ratios (%) for 45 min PH using the OhioT1DM dataset (Best method is shown in bold.)

Method	None (0)	Slight (1)	Moderate (2)	Great (3)	Extreme (4)
LSTM	80.95	18.98	0.07	0.00	0.00
WaveNet	79.88	20.07	0.05	0.00	0.00
GRU	81.08	18.85	0.07	0.00	0.00
WaveNet + LSTM	81.04	18.90	0.06	0.00	0.00
WaveNet + GRU	81.01	18.92	0.07	0.00	0.00
LSTM + GRU	81.10	18.83	0.07	0.00	0.00
LSTM + WaveNet + GRU	**81.14**	**18.80**	**0.06**	**0.00**	**0.00**

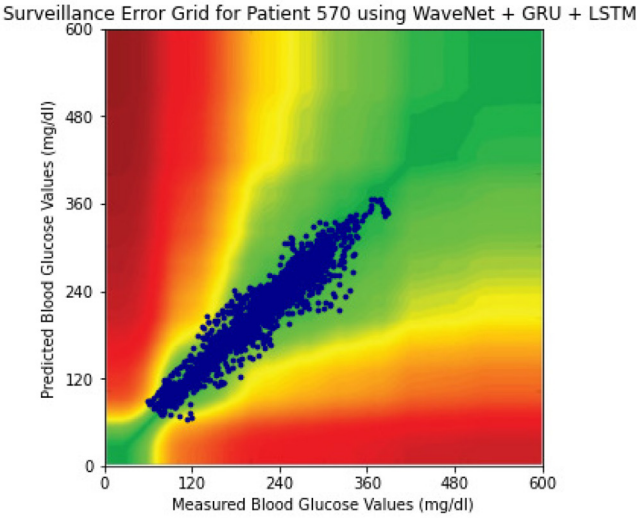
Table 11 – SEG average risk range distribution ratios (%) for 60 min PH using the OhioT1DM dataset (Best method is shown in bold.)

Method	None (0)	Slight (1)	Moderate (2)	Great (3)	Extreme (4)
LSTM	76.23	23.63	0.14	0.00	0.00
WaveNet	75.52	24.37	0.11	0.00	0.00
GRU	75.92	23.95	0.13	0.00	0.00
WaveNet + LSTM	76.21	23.67	0.12	0.00	0.00
WaveNet + GRU	76.04	23.84	0.12	0.00	0.00
LSTM + GRU	**76.26**	**23.59**	**0.15**	**0.00**	**0.00**
LSTM + WaveNet + GRU	76.13	23.74	0.13	0.00	0.00

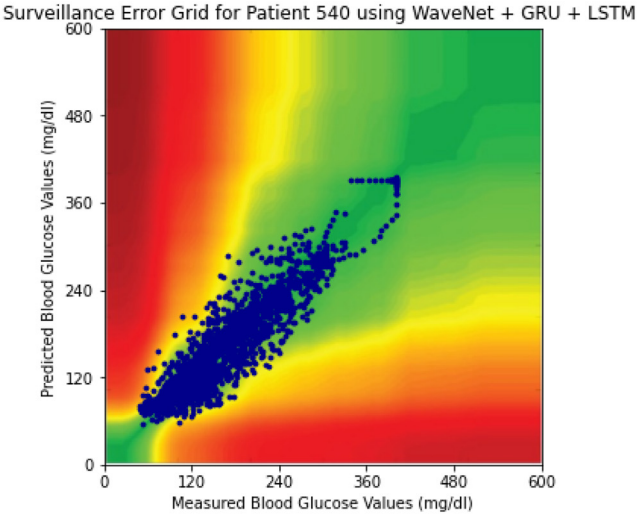
using 30 min patient BG history, and then BG prediction was performed for 30 min, 45 min, and 60 min PH. As seen in Fig. 2, the predictions for the three models are also passed through a rule-based decision to create decision level fusion of LSTM and GRU models, LSTM and WaveNet models, GRU and WaveNet models, and finally the fusion of LSTM, GRU and WaveNet models. Thus, seven different results were obtained for BG prediction for every patient. Detailed information on the proposed weighted decision level fusion in this paper is given in subSection 2.2.

In experimental studies, 12 patients are tested separately and the average evaluation results are given in Tables 5–7 for predictions of 30 min, 45 min and 60 min, respectively. When the tables are examined, the lowest RMSE values were obtained by "LSTM + WaveNet + GRU", for all PHs. When all evaluation metrics are considered, it was observed that the most successful method for BG prediction is the fusion of three networks. In addition, in the Table 8, the results of four different numerical evaluation metrics are given to show how the fusion method proposed in the paper works for each patient in the OhioT1DM dataset. When the Table 8 is examined, it is seen that the lowest RMSE value was obtained with patient ID 570, and the highest RMSE value was obtained with patient ID 540. While performing the risk assessment with the surveillance error grid, risk classes are ranked between 0 and 4 according to the difference between the predicted value and the measured value. None (0) indicates errors with the lowest risk, extreme (4) indicates errors with the highest and dangerous risks. In Table 9–11 the SEG average risk range distribution ratios for seven different BG prediction methods are given for 30 min, 45 min and 60 min PH. When SEG average risk distribution tables were examined, it has been observed that the most successful method for short-term BG prediction is "LSTM + WaveNet + GRU", while "LSTM + GRU" shows the best results for 60 min PH.

When the BG prediction results of the patients in the dataset were examined with the proposed methods, the most successful prediction results were obtained for the patient with ID 570. The lowest RMSE values obtained for 30 min, 45 min, and 60 min PH for this patient are 18.06 (with "LSTM + WaveNet + GRU"), 23.88 (with GRU), and 28.86 (with LSTM) respectively. The SEGs for this patient's 30 min PH BG predictions obtained with the seven proposed methods are given in Fig. 9. The reason for plotting SEG for BG prediction is to ensure that the error in the prediction performed is not in a way that endangers the patient's life. These areas on the grid are shown in red. When the SEGs given in Fig. 9 patient ID 570 (best performance) and patient ID 540 (lowest performance) for 30 min PH were examined, it was observed that the predictions made by the proposed method were compatible with the

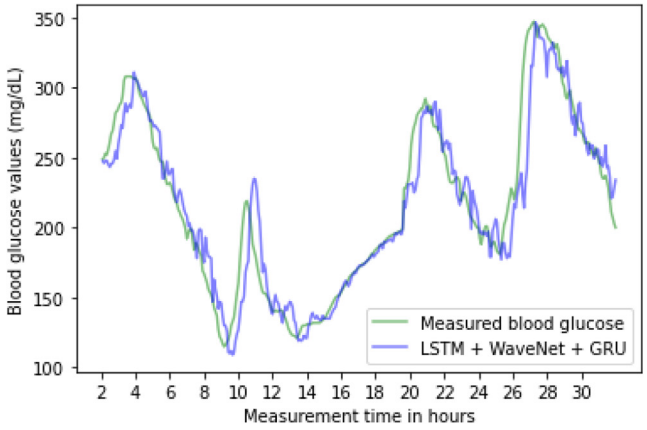


(a)

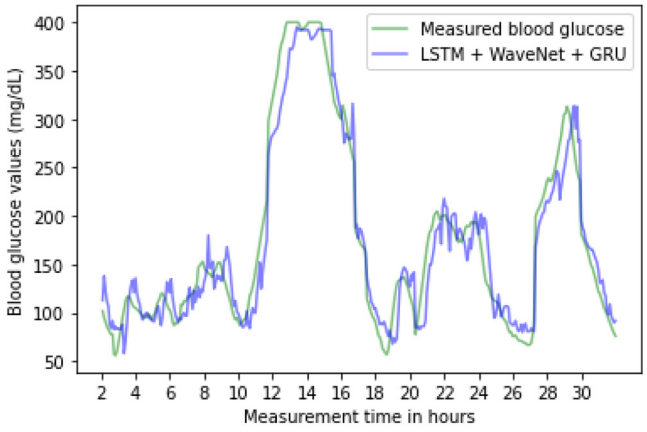


(b)

Fig. 9 – SEGs for (a) patient ID 570 (best performance) and (b) patient ID 540 (lowest performance) with 30 min PH BG predictions obtained with the proposed LSTM + WaveNet + GRU fusion method.



(a)



(b)

Fig. 10 – Comparison of measured blood glucose and prediction results obtained with the proposed LSTM + WaveNet + GRU fusion method for (a) patient ID 570 (best performance) and (b) patient ID 540 (lowest performance) for 30 min PH.

measured BG levels. To show this compatibility between the measured and the predicted data, the graphics plotted using the test data of patient ID 570 (best performance) and patient ID 540 (lowest performance) are given in Fig. 10 for 30 min PH. When the measured and predicted blood glucose values shown in Fig. 10 are examined, it is seen that the predictions give results close to the measured values generally, however at the peak values, predictions do not follow the measured blood glucose as closely. Considering the PRED-EGA values (Table 12) in which the prediction success of the proposed method is evaluated, it has been observed that the data in normoglycemia and hyperglycemia levels of the fusion method gives successful results in prediction, but cannot show the same success in hypoglycemia levels. When the reason for this low performance was examined, it was seen that it was caused by the low number of data at the level of hypoglycemia.

As a result of the experimental studies, when the numerical evaluation metrics and grids obtained for networks trained using 30 min history were examined, it was observed that the results did not exceed the low-risk regions and when compared with other studies in the literature, RMSE values were close to the average of state-of-art studies, although a lower time window was used in the training of networks in this paper.

5. Conclusion and discussion

In this paper, a method in which three different deep learning networks were used and weighted decision level fusions were proposed for BG prediction. The patient history was selected as 30 min measurement for the training of three different deep neural networks in the proposed method. Although three different deep learning networks mentioned in the literature are used one by one for BG prediction, in this study new BG prediction has been realized by fusion of the prediction values obtained from these three networks. When the experimental studies conducted with the OhioT1DM dataset were examined, it was observed that for all selected prediction

Table 12 – PRED-EGA table for 30 min PH using the OhioT1DM dataset performed with the proposed fusion method.

	BG ⩽70 mg/dL				BG 70–180 mg/dL				BG ⩾180 mg/dL			
Patient ID	Accurate (%)	Benign (%)	Error (%)	# of data	Accurate (%)	Benign (%)	Error (%)	# of data	Accurate (%)	Benign (%)	Error (%)	# of data
540	19.35	3.23	77.42	124	76.98	18.86	4.16	1442	75.95	16.84	7.22	790
544	54.84	0.00	45.16	31	89.25	8.78	1.97	1780	77.53	17.53	4.94	812
552	29.69	1.56	68.75	64	86.24	11.66	2.09	1432	84.29	12.86	2.86	351
559	22.22	0.00	77.78	72	90.42	7.00	2.58	1358	79.87	16.29	3.83	939
563	40.00	0.00	60.00	15	85.96	10.39	3.66	1533	79.77	16.63	3.59	974
567	61.25	1.88	36.88	160	84.41	12.70	2.88	1283	74.94	17.38	7.67	444
570	9.09	9.09	81.82	11	93.92	5.20	0.89	789	87.55	10.86	1.59	1826
575	41.01	3.60	55.40	139	87.24	8.53	4.23	1586	79.05	12.38	8.57	734
584	0.00	4.00	96.00	25	83.20	11.96	4.84	1530	81.46	12.43	6.11	917
588	0.00	0.00	100.00	4	85.22	11.17	3.61	1495	83.17	13.88	2.95	1255
591	9.56	5.88	84.56	136	85.44	11.06	3.50	1829	82.47	13.32	4.21	736
596	20.27	6.76	72.97	74	89.67	8.13	2.20	1957	80.10	16.56	3.34	629
Average	**25.61**	**3.00**	**71.39**	**71**	**86.50**	**10.45**	**3.05**	**1501**	**80.51**	**14.75**	**4.74**	**867**

horizons BG prediction "LSTM + WaveNet + GRU" fusion performance gave more successful results and this is the significance of the study. RMSE value of the three network fusion, which gives the best results in the test process, is 21.90 mg/dl, 29.12 mg/dl, and 31.10 mg/dl for 30 min, 45 min, and 60 min PH respectively. The study also provided SEGs for patient ID 570, the patient with the lowest RMSE value. These seven grids clearly show that none of the prediction data performed with the proposed methods were in the high-risk regions.

In order to compare the studies in the literature with the proposed method, studies that conducted experimental studies using the OhioT1DM dataset were examined. Studies using this dataset have generally performed their experimental studies using the first version of the OhioT1DM dataset containing 6 patient data. In the study conducted by Martinsson et al. [6] in 2018, by using the first version of the OhioT1DM dataset and using a 30 min patient history with LSTM networks, the RMSE value of the naive baseline of predicting the last value was obtained as 22.5 mg/dl and 36.6 mg/dl for 30 and 60 min PHs. In another study conducted by Martinsson et al. [13] in 2019, with the same dataset using 60 min patient history and LSTM networks, by naively predicting the last value the RMSE value was obtained as 22.43 mg/dl and 36.39 mg/dl for 30 and 60 min PHs respectively. In 2020, Li et al. [16] by using the first version of the OhioT1DM dataset, experimental studies were carried out for five different methods. In this study, as a result of experimental studies conducted with networks trained using 80 min history, RMSE values varying between 19.28 mg/dl and 22.93 mg/dl for 30 min PH were obtained, while for 60 min PH RMSE values varying between 31.83 mg/dl and 39.53 mg/dl were obtained. A benchmarking study was conducted by Xie and Wang in late 2020, using a total of 11 machine learning, two of which are deep learning networks, on the first version of the OhioT1DM dataset [17]. The RMSE values during the day obtained as a result of this study vary between 21.81 mg/dl and 25.44 mg/dl. Detailed information on the studies carried out in the last three years on BG prediction using The OhioT1DM dataset in the literature is given in Table 13. When these studies are compared with the proposed fusion method, considering the data types and history used in experimental studies, it has been shown that the proposed method works successfully in terms of RMSE value for BG prediction. Compared to other studies in the literature, the use of a larger dataset in this study, and the increase in RMSE as a result of the BG prediction made with the proposed fusion method, is very important for a disease with important and fatal consequences such as diabetes. Due to the absence of a similar hybrid method in the literature, the proposed fusion method offers a new perspective for future studies for BG prediction in the literature. When the proposed fusion method is used for BG prediction with 30-min PH, which is necessary to prevent potentially dangerous glycemic attacks in the future, the computational cost brought to the system during prediction after model training is negligible compared to the increase in prediction accuracy.

In future studies, instead of making BG predictions using only the patient's own history, BG prediction will be attempted with a common model created by using the history of many diabetes patients. The fusion method proposed in this study is also planned to be integrated into the future study. The weighted decision level fusion coefficients in this paper were determined for this pioneering study using data from 12 patients. In future studies, when the dataset is expanded, these coefficients should be updated by performing the necessary exploratory data analysis for these new data. In addition to the model development studies, it is planned to use the prediction values made using the fusion model, as in Sahin and Aydin's study in 2021 [24], in calculating the values required for insulin injection. If these works are successful, it is also among the plans to make the developed system a mobile application, as in the Naumova et al.'s study [37].

Table 13 – Comparison of blood glucose prediction studies using OhioT1DM data set (2018–2021) (BG denotes Blood glucose level, IOI denotes Injection of insulin, M denotes Meal, CI denotes Carbohydrate intake, HR denotes Heart rate, PA denotes Physical activity.).

	Dataset	# of Patients	History	PH	RMSE	Data Types
Martinsson et al. (2018) [6]	2018	6	30 min	30 min	22.50	BG
	2018	6	30 min	60 min	36.60	BG
Li et al. [16]	2018	6	80 min	30 min	19.28	BG + IOI + M
	2018	6	80 min	60 min	31.83	BG + IOI + M
De Bois et al. [22]	2018	6	3-h	30 min	20.01	BG
	2018	6	3-h	60 min	31.97	BG
Xie and Wang [17]	2018	6	60 min	30 min	19.48	BG + IOI + CI + HR
Martinsson et al. (2020) [13]	2018	6	60 min	30 min	22.43	BG
	2018	6	60 min	60 min	36.38	BG
Sahin and Aydin [24]	2018	6	25 min	30 min	18.81	BG + IOI + CI + PA
	2018	6	25 min	60 min	30.89	BG + IOI + CI + PA
LSTM + WaveNet + GRU (proposed method)	**2018 + 2020**	**12**	**30 min**	**30 min**	**21.90**	**BG**
	2018 + 2020	**12**	**30 min**	**60 min**	**35.10**	**BG**

CRediT authorship contribution statement

Hatice Vildan Dudukcu: Conceptualization, Methodology, Software, Writing - original draft. **Murat Taskiran:** Conceptualization, Methodology, Software, Writing - original draft. **Tulay Yildirim:** Conceptualization, Methodology, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] Pouya Saeedi, Paraskevi Salpea, Suvi Karuranga, Inga Petersohn, Belma Malanda, Edward W Gregg, Nigel Unwin, Sarah H Wild, Rhys Williams. Mortality attributable to diabetes in 20–79 years old adults, 2019 estimates: Results from the international diabetes federation diabetes atlas. Diabetes research and clinical practice; 2020. p. 108086.

[2] Ashenafi Zebene Woldaregay, Eirik Ørsand, Ståle Walderhaug, David Albers, Lena Mamykina, Taxiarchis Botsis, Gunnar Hartvigsen. Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. Artif Intell Med 2019;98:109–134.

[3] Oviedo Silvia, Vehí Josep, Calm Remei, Armengol Joaquim. A review of personalized blood glucose prediction strategies for t1dm patients. Int J Numer Methods Biomed Eng 2017;33(6) e2833.

[4] J Ignacio Hidalgo, J Manuel Colmenar, Gabriel Kronberger, Stephan M Winkler, Oscar Garnica, Juan Lanchares. Data based prediction of blood glucose concentrations using evolutionary methods. J Med Syst 2017;41(9):142.

[5] Doike Takuyoshi, Hayashi Kenya, Arata Shigeki, Mohammad Karim Nissar, Kobayashi Atsuki, Niitsu Kiichi. A blood glucose level prediction system using machine learning based on recurrent neural network for hypoglycemia prevention. In: 2018 16th IEEE International New Circuits and Systems Conference (NEWCAS), IEEE. p. 291–5.

[6] John Martinsson, Alexander Schliep, Bjorn Eliasson, Christian Meijner, Simon Persson, Olof Mogren. Automatic blood glucose prediction with confidence using recurrent neural networks. In: KHD@ IJCAI, 2018.

[7] Taiyu Zhu, Kezhi Li, Pau Herrero, Jianwei Chen, and Pantelis Georgiou. A deep learning algorithm for personalized blood glucose prediction. In: KHD@ IJCAI, 2018, p. 64–78.

[8] Takoua Hamdi, Jaouher Ben Ali, Véronique Di Costanzo, Farhat Fnaiech, Eric Moreau, and Jean-Marc Ginoux. Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm. Biocybern Biomed Eng 2018;38(2):362–372.

[9] Jaouher Ben Ali, Takoua Hamdi, Nader Fnaiech, Véronique Di Costanzo, Farhat Fnaiech, Jean-Marc Ginoux. Continuous blood glucose level prediction of type 1 diabetes based on artificial neural network. Biocybern Biomed Eng 2018;38 (4):828–840.

[10] Aliberti Alessandro, Pupillo Irene, Terna Stefano, Macii Enrico, Di Cataldo Santa, Patti Edoardo, Acquaviva Andrea. A multi-patient data-driven approach to blood glucose prediction. IEEE Access 2019;7:69311–25.

[11] Yang Jun, Li Lei, Shi Yimeng, Xie Xiaolei. An arima model with adaptive orders for predicting blood glucose concentrations and hypoglycemia. IEEE J Biomed Health Inf 2018;23(3):1251–60.

[12] Wei Song, Wanyuan Cai, Jing Li, Fusong Jiang, Shengqi He. Predicting blood glucose levels with emd and lstm based cgm data. In: 2019 6th International Conference on Systems and Informatics (ICSAI), IEEE; 2019, p. 1443–1448.

[13] Martinsson John, Schliep Alexander, Eliasson Björn, Mogren Olof. Blood glucose prediction with variance estimation using recurrent neural networks. J Healthcare Inf Res 2020;4 (1):1–18.

[14] Xia Yu, Rashid Mudassir, Feng Jianyuan, Hobbs Nicole, Hajizadeh Iman, Samadi Sediqeh, et al. Online glucose prediction using computationally efficient sparse kernel filtering algorithms in type-1 diabetes. IEEE Trans Control Syst Technol 2018.

[15] Li Kezhi, Daniels John, Liu Chengyuan, Herrero Pau, Georgiou Pantelis. Convolutional recurrent neural networks for glucose prediction. IEEE J Biomed Health Inf 2019;24 (2):603–13.

[16] Li Kezhi, Liu Chengyuan, Zhu Taiyu, Herrero Pau, Georgiou Pantelis. Glunet: a deep learning framework for accurate glucose forecasting. IEEE J Biomed Health Inf 2019;24 (2):414–23.

[17] Xie Jinyu, Wang Qian. Benchmarking machine learning algorithms on blood glucose prediction for type 1 diabetes in comparison with classical time-series models. IEEE Trans Biomed Eng 2020.

[18] Wang Wenbo, Tong Meng, Min Yu. Blood glucose prediction with vmd and lstm optimized by improved particle swarm optimization. IEEE Access 2020.

[19] Zhu Taiyu, Yao Xi, Li Kezhi, Herrero Pau, Georgiou Pantelis. Blood glucose prediction for type 1 diabetes using generative adversarial networks. CEUR Workshop Proc 2020;2675:90–4.

[20] Ganjar Alfian, Muhammad Syafrudin, Jongtae Rhee, Muhammad Anshari, M Mustakim, Imam Fahrurrozi. Blood glucose prediction model for type 1 diabetes based on extreme gradient boosting. In IOP Conference Series: Materials Science and Engineering, vol. 803, IOP Publishing, 2020. p. 012012.

[21] Cindy Marling and Razvan C Bunescu. The ohiot1dm dataset for blood glucose level prediction. In: KHD@ IJCAI, 2018. p. 60–63.

[22] Maxime De Bois, Mehdi Ammi, Mounīm A El Yacoubi. Glyfe: Review and benchmark of personalized glucose predictive models in type-1 diabetes. arXiv preprint arXiv:2006.15946, 2020.

[23] Md Fazle Rabby, Yazhou Tu, Md Imran Hossen, Insup Lee, Anthony S Maida, Xiali Hei. Stacked lstm based deep recurrent neural network with kalman smoothing for blood glucose prediction. BMC Med Inf Decis Mak 2021;21 (1):1–15.

[24] Asiye Şahin, Ahmet Aydın. Personalized advanced time blood glucose level prediction. Arab J Sci Eng 2021 p. 1–12.

[25] Cindy Marling, Razvan Bunescu. The ohiot1dm dataset for blood glucose level prediction: Update 2020. KHD@ IJCAI, 2020.

[26] Gábor Petneházi. Recurrent neural networks for time series forecasting. arXiv preprint arXiv:1901.00069, 2019.

[27] Hochreiter Sepp, Bengio Yoshua, Frasconi Paolo, Schmidhuber Jürgen, et al., editors. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.

[28] Hochreiter Sepp, Schmidhuber Jürgen. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[29] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078, 2014.
[30] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu. Wavenet: A generative model for raw audio. arXiv preprint arXiv:1609.03499, 2016.
[31] LaValle Steven M, Branicky Michael S, Lindemann Stephen R. On the relationship between classical grid search and probabilistic roadmaps. Int J Robot Res 2004;23(7–8):673–92.
[32] Chai Tianfeng, Draxler Roland R. Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. Geosci Model Develop 2014;7 (3):1247–50.
[33] Ummul Khair, Hasanul Fahmi, Sarudin Al Hakim, Robbi Rahim. Forecasting error calculation with mean absolute deviation and mean absolute percentage error. In: Journal of Physics: Conference Series, vol. 930, IOP Publishing, 2017, p. 012002.
[34] David C Klonoff, Courtney Lias, Robert Vigersky, William Clarke, Joan Lee Parkes, David B Sacks, M Sue Kirkman, Boris Kovatchev, Error Grid Panel. The surveillance error grid. J Diabet Sci Technol 2014;8(4):658–672.
[35] Kovatchev Boris P, Wakeman Christian A, Breton Marc D, Kost Gerald J, Louie Richard F, Tran Nam K, Klonoff David C. Computing the surveillance error grid analysis: procedure and examples. J Diabet Sci Technol 2014;8(4):673–84.
[36] Sampath Sivananthan, Valeriya Naumova, Chiara Dalla Man, Andrea Facchinetti, Eric Renard, Claudio Cobelli, Sergei V Pereverzyev. Assessment of blood glucose predictors: the prediction-error grid analysis. Diabet Technol Therap 2011;13 (8):787–796.
[37] Naumova Valeriya, Nita Lucian, Poulsen Jens Ulrik, Pereverzyev Sergei V. Meta-learning based blood glucose predictor for diabetic smartphone app. In: Prediction methods for blood glucose concentration. Springer; 2016. p. 93–105.