Glucose Transformer: Forecasting Glucose Level and Events of Hyperglycemia and Hypoglycemia

Sang-Min Lee, Dae-Yeon Kim, and Jiyoung Woo

***Abstract*—To avoid the adverse consequences from abrupt increases in blood glucose, diabetic inpatients should be closely monitored. Using blood glucose data from type 2 diabetes patients, we propose a deep learning model-based framework to forecast blood glucose levels. We used continuous glucose monitoring (CGM) data collected from inpatients with type 2 diabetes for a week. We adopted the Transformer model, commonly used in sequence data, to forecast the blood glucose level over time and detect hyperglycemia and hypoglycemia in advance. We expected the attention mechanism in Transformer to reveal a hint of hyperglycemia and hypoglycemia, and performed a comparative study to determine whether Transformer was effective in the classification and regression of glucose. Hyperglycemia and hypoglycemia rarely occur and this results in an imbalance in the classification. We built a data augmentation model using the generative adversarial network. Our contributions are as follows. First, we developed a deep learning framework utilizing the encoder part of Transformer to perform the regression and classification under a unified framework. Second, we adopted a data augmentation model using the generative adversarial network suitable for time-series data to solve the data imbalance problem and to improve performance. Third, we collected data for type 2 diabetic inpatients for mid-time. Finally, we incorporated transfer learning to improve the performance of regression and classification.**

***Index Terms*—Type 2 diabetes, continuous glucose monitoring, hyperglycemia, hypoglycemia, deep learning, Transformer.**

I. Introduction

DIABETES mellitus is a type of metabolic disorder marked by excessive blood glucose levels and characterized by insufficient insulin secretion or impaired normal function [1]. Serious problems such as blurred vision, heart failure, and cerebral infarction can arise if diabetes is left untreated for a long time. According to the International Diabetes Federation (IDF), approximately 537 million individuals (aged 20–79) will have diabetes by 2021, with the number of persons affected expected to rise [2]. Diabetes can be classified into three types, not only type 1 and type 2 diabetes but also gestational diabetes. Among these, type 2 diabetes is characterized by a decrease in insulin function due to insulin resistance. As a result, in some type 2 diabetes patients, the pancreas depletes insulin and produces less insulin, resulting in hyperglycemia (high blood glucose levels). Type 2 diabetes is the most common type of diabetes, accounting for approximately 90% of all cases. It is most common in the elderly but has recently become more common in children and young adults. When diagnosed with type 2 diabetes, blood glucose levels can be maintained with diet, exercise, and medication including oral or injectable hypoglycemic agents and insulin. Blood glucose levels should be continually monitored and managed in hospitalized patients. The doctor can help stabilize the patient's blood glucose level in case of abrupt variations during the monitoring process. It is difficult for medical staff to monitor each patient because the amount of medical staff is minimal compared to the number of diabetic patients hospitalized. Furthermore, the time it takes for medical personnel to come and stabilize the patient's blood glucose level varies, and the patient's condition may worsen with time. As a result, this study aims to create a deep learning model that predicts future blood glucose levels for patients with type 2 diabetes who are hospitalized so that if blood glucose variations are expected, medical staff can be notified ahead of time. For this type 2 diabetes study, we used data obtained from patients with type 2 diabetes who had been admitted to the hospital using a continuous glucose monitoring (CGM) device. CGM is a commercially available device, and it is already widely used in type 1 diabetes patients and type 2 diabetes patients. In addition, according to the recent guidelines for diabetes treatment, it is recommended to use blood glucose information obtained through CGM to assess glycemic control. The CGM device is one of the monitoring methods for diabetes patients' blood glucose control, and it identifies blood glucose variations or hypoglycemia efficiently. The device eliminates the pain and anxiety associated with drawing blood with a finger stick and the associated time and cost. With the device, blood glucose fluctuations can be monitored in real-time during all food intake and physical activity, allowing for prompt response in an emergency [3]. CGM consists of a sensor that measures

Manuscript received 29 July 2022; revised 15 November 2022; accepted 24 December 2022. Date of publication 2 February 2023; date of current version 7 March 2023. This work was supported in part by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean Government (MSIT) under Grant 2019M3E5D1A02069069, in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2020R1I1A3056858, and in part by Soonchunhyang University Research Fund. *(Corresponding authors: Dae-Yeon Kim; Jiyoung Woo.)*

Sang-Min Lee is with the Department of ICT Convergence, Soonchunhyang University, Asan 31538, South Korea (e-mail: dsp12357@sch.ac.kr).

Dae-Yeon Kim is with the Department of Internal Medicine, Soonchunhyang University Cheonan Hospital, Cheonan 31151, South Korea (e-mail: c99851@schmc.ac.kr).

Jiyoung Woo is with the Department of Big Data Engineering, Soonchunhyang University, Asan 31538, South Korea, and also with the Management Information System Department of Arizona University, Tucson, AZ 85721 USA (e-mail: jywoo@sch.ac.kr, jywoo@arizona.edu).

Digital Object Identifier 10.1109/JBHI.2023.3236822

glucose for seven days and a transmitter that wirelessly shares glucose information. The transmitter has a battery life of 90 days and is reusable.

In this paper, we propose a deep learning framework that uses the encoder part of Transformer [4] incorporated with data augmentation using generative learning to predict and classify the blood glucose levels of type 2 diabetes patients obtained using a CGM device. Transformer was created for natural language processing. It performs well in image processing and in language translation [5], as its fields of application have expanded. Furthermore, Transformer compensates for performance deterioration by building a model using the self-attention technique. Thus, it solves the performance degradation that occurs as the input length increases, which is a drawback of the recurrent neural network (RNN) [6]. Unlike RNN-based models, which cannot be processed in parallel owing to their sequential nature, the attention of the Transformer model offers the advantage of stacking numerous layers in parallel and hierarchically to boost speed and performance. We do not use patient identification or blood test information from the acquired data; instead, we use only blood glucose levels. Our work makes the following contributions. First, we propose using the Transformer and especially the encoder part for performing both time-series forecasting and classification tasks using an identical framework. We expect that the Transformer's attention will hint at the occurrence of rare events. Furthermore, positional encoding will be helpful in processing data sequentially, compensating for RNN-based models. Second, we incorporate generative learning for time-series forecasting to resolve the data imbalance problem. The long short-term memory (LSTM)-based generative adversarial network (GAN) model generates synthetic glucose time series. Third, we collect the inpatient data on type 2 diabetes for up to a week. In particular, mid-to-long term continuous glucose data of type 2 diabetes inpatients are difficult to find in the literature, even though type 2 diabetes should be carefully monitored, especially in hospitals. Finally, we adopt transfer learning for data augmentation using the open dataset of type 1 diabetes [7].

II. Related Work

The number of studies using machine learning and deep learning to use data acquired from hospital inpatients and surgical patients has expanded owing to the recent developments in artificial intelligence. Many studies have been conducted recently on blood glucose categorization and prediction, and statistical and deep learning models have been used to predict blood glucose levels, which are time series data. The auto-regressive integrated moving average (ARIMA) model [8], which is an example of a statistical model, is a form of forecasting that provides data with linearity and regularity. However, because the blood glucose data collected from patients is irregular and non-linear, it may be challenging to obtain good performance using the ARIMA model [9]. Deep learning model experiments have been carried out recently to overcome this challenge. Deep learning has the advantage of automatically extracting features from input data, supporting multiple inputs and outputs, and finding patterns from sequences with relatively large input lengths. RNN-based models, unlike deep neural network (DNN) [10] and convolutional neural network (CNN) [11] models, do particularly well on sequence data. The RNN retains and learns the last time's data within, and it can recognize a pattern distributed over a lengthy period. However, as the input text grows longer, the gradient vanishing problem arises, in which earlier information is lost. LSTM [12] and gated recurrent unit (GRU) [13] sharing cell states have surfaced to overcome this problem. However, because the meaning of all input data must be encoded in a fixed-length hidden state vector, there is an information loss problem and persistent gradient vanishing problem. The attention model was presented to overcome the problem mentioned above. In sequence to sequence (Seq2Seq) [14], which is an RNN-based encoder-decoder model, the attention model is used to overcome the performance decrease due to information loss that occurs when the input sequence becomes longer. The attention model consults the whole input sequence from the encoder at every place when the Seq2Seq decoder predicts the output value, focusing mainly on the part of the input data linked to the value to be predicted [15]. In this way, the attention model that handles the problem of information loss and gradient loss performs well and has been applied in various sectors.

Even though type 2 diabetes accounts for most diabetic patients in blood glucose-related studies, the research on type 2 has been left behind because it has a high level of variability in blood glucose, making it hard to anticipate the value. Therefore, type 1 diabetes has been the most frequently studied. Tomas and Michael [16] proposed a new low-complexity, explainable blood glucose prediction method derived from the rule-based model rather than a complex deep learning model. They defined several glucose change patterns, extracted features, and then experimented with predicting the patterns. They evaluated using the Ohio data set and the surveillance error grid (SEG). When predicting blood glucose levels after 30 min, 96.3% achieved a relative error of less than 30%. This work has the drawback that the relative error was evaluated as 0 when the blood glucose level exceeded the set upper/lower limit. Nemat et al. [17] used deep learning and ensemble models to predict blood glucose levels after 30 and 60 min to manage type 1 diabetes more effectively. The OhioT1DM dataset, which included 12 patients diagnosed with diabetes, was used. The three types of models, including linear regression, vanilla LSTM (VLSTM), and bidirectional LSTM (BiLSTM) [18] were tested in the experiment, and the three models were ensembled using the stacking, multivariate, and sub-sequences approaches. The ensemble model outperformed the existing models, and the stacking method outperformed all. Daniels et al. [19] used multi-task learning to create a personalized blood glucose prediction model. To learn many tasks simultaneously utilizing various output layers in a single neural network, a multi-task learning technique uses the associations between tasks. They addressed the issue of individual blood glucose variability through this method. The OhioT1DM data set, which includes 12 persons with type 1 diabetes, was the source of information used in the experiment. To perform single-task learning, transfer learning, and multi-task learning, the convolutional recurrent neural network (CRNN) model was employed, and the performance was compared to

support vector regression (SVR), a machine learning model. Additionally, multi-task learning was performed by splitting the experiment into two groups based on blood glucose variability. The prediction period was divided into segments ranging in length from 30 to 120 min for the experiment, and 30 min produced the best results. The root mean square error (RMSE) of 18.8 ± 2.3 and mean absolute error (MAE) of 13.2 ± 1.6 from multi-task learning showed impressive performance over other models. Asad et al. [20] aimed to forecast blood glucose levels overcoming the insulin absorption delay time of type 1 diabetes. They used AIDA to simulate blood glucose data and real CGM data to compare the feed-forward and the nonlinear auto-regressive neural networks. The anticipated time was set to 15 min, and 30 min to compare the results. The RMSE was utilized as a performance indicator. The average RMSEs for ideal feed-forward neural networks were 1.21 (mg/dL) and 3.64 (mg/dL), respectively, while the average RMSEs for optimal nonlinear autoregressive neural networks were 0.695 (mg/dL) and 1.39 (mg/dL). Kim et al. [21] conducted a study to predict blood glucose levels by collecting CGM data from 21 hospitalized type 2 diabetes patients and using an RNN-based deep learning model. The model was set to forecast the blood glucose value 30 min after the input by setting the delay time to 30 min. The RNN, GRU, and LSTM models were used to develop a personalized model that used an individual's previous data for prediction and builds the model for each patient. A mean RMSE of 21.5 (mg/dL) and a mean absolute percentage error (MAPE) of 11.1 percent were attained in 21 hospitalized individuals. Owing to the small number of samples, this personalized model for an individual patient was built, but the need for the individual's data to be applicable to a particular patient has poor applicability. Mario [22] built a physiological prediction model to help people with type 1 diabetes, which inhibits the pancreas' ability to generate insulin. The AIDA simulator was used to produce type 1 diabetes patient data, and the D1NAMO open data set containing actual patient information was employed in the experiment. With CGM data, carbohydrate digestion, and insulin absorption processes as inputs, an LSTM was used to predict glucose levels between 30 and 60 min ahead. RMSEs of 6.42 (mg/dL) and 11.35 (mg/dL) in the actual patient data were achieved in the estimation of blood glucose levels after 30 and 60 min, respectively, and 3.45 (mg/dL) in the simulator generated data after 30 min. Zhu et al. [23] used the dilated recurrent neural network (DRNN) [24] model to predict future glucose levels 30 min ahead. The DRNN was created to overcome the difficulties of vanishing/exploding gradients and efficient parallelization that RNNs face when dealing with longer sequences. The hidden state before the skip length is reflected in this model, not the method of reflecting the initial hidden state to the RNN. The skip length of the DRNN was set at [1,2,4] in the previous experiment, and the results were compared to those of the auto-regressive model, SVR, and conventional neural networks for glucose prediction. To assess the performance of each model, the OhioT1DM data were used. The autoregressive model, SVR, and conventional neural networks for glucose prediction models, achieved 20.1, 21.7, and 22.9 (mg/dL) RMSEs, respectively, whereas the DRNN model achieved 18.9 (mg/dL) RMSE, showing better performance. Li et al. [25] proposed a CRNN model as a deep learning model for type 1 diabetes patients to predict blood glucose after 30 and 60 min. A 1D Gaussian kernel filter was included in the convolution layer, and the RNN model used to learn long-term dependencies was LSTM. The experiment used ten patients' data and clinical data produced using UVA/Padova type 1 diabetes. Real patient data and generated patient data were used in the experiment. Both experiments performed well when the prediction horizon was 30 min, with the virtual patient data recording RMSE of 9.38 and mean absolute relative difference (MARD) of 5.50, and the real patient data recording RMSE of 21.07 and MARD of 11.61. Deng et al. [26] used the OhioT1DM data set to construct a deep learning-based predictive model to classify future blood glucose levels into hypoglycemia, hyperglycemia, and normoglycemia between 5 min and 1 h. The problem was solved using transfer learning and data augmentation because there was lack of data for model training and the rate of hypoglycemia was low. Seo et al. [27] employed many machine learning techniques to predict postprandial hypoglycemia. Data from 52 patients with type 1 diabetes and 52 patients with type 2 diabetes, who were adult patients over 18, were collected for three days. Postprandial hypoglycemia was predicted using random forest (RF), support vector machine (SVM), K-nearest neighbor (KNN), and linear regression (LR) models after 30 min. With an average area under the curve (AUC) of 0.966, an average sensitivity of 89.6%, an average specificity of 91.3%, and an average F1 score of 0.543, the RF model outperformed the others. Bertachi et al. [28] developed an individualized machine learning model for predicting nocturnal hypoglycemia in patients with type 1 diabetes using CGM and physical activity monitors. The data used in the experiment were collected from 10 type 1 diabetes patients at home through a wearable device. A multilayer perceptron (MLP) and SVM were used as personalized prediction models, and k-fold cross-validation was performed. The results of the SVM showed improved performance compared to those of the MLP, with a sensitivity of 78%, specificity of 82%, and accuracy of 80%. Jensen et al. [29] collected CGM, diet, insulin, and statistical data from 463 patients with type 1 diabetes to predict nocturnal hypoglycemia. Because it occurs during sleep, nocturnal hypoglycemia is difficult to detect. This study defined nocturnal hypoglycemia as three consecutive CGM readings of less than 54 (mg/dL). Hypoglycemia and normoglycemia were classified using the linear discriminant analysis classifier model. The experimental results achieved a sensitivity of 75% and specificity of 70%. Mordvanyuk et al. [30] used data generated by the UVA-Padova type 1 diabetes simulator, including 346 hypoglycemic and 2,833 hyperglycemic cases to predict whether they would fall into an abnormal state between 2 and 6 h later. Time, carbohydrate intake, insulin dose, and CGM readings parameters were used to create events for each meal. In the experimental results, 88% sensitivity in hypoglycemia and 82% sensitivity in hyperglycemia was achieved using the KNN algorithm only for carbohydrate intake, volume dose, and pre-meal blood glucose information when using sequence data. They earned good performance because they used the simulated dataset, which comprised finely constructed data, rather than

CGM and predicted the event during a period rather than a specific time point. The prediction on an exact time point is much more difficult than the prediction during a long period.

From the literature review, we found that most previous works deal with the type 1 diabetes, which occurs mainly in children. For the type 2 diabetes, especially inpatients, the glucose level should be monitored to lower the risk of hyperglycemia or hypoglycemia occurrence. The regression for predicting the future glucose levels has been intensely studied, but there is a lack of studies on the classification of hyperglycemia or hypoglycemia. The main objective of the regression is to identify events that could endanger the patient, and thus the classification should also be addressed. More importantly, the previous works on classification set their goals to predict the event during a period rather than at a specific time point. The studies aimed at predicting the event during a period exhibited better performance, but it is necessary to point out the exact event time to prepare and inject the insulin to relieve the hyperglycemia or hypoglycemia. The model to specify the event time showed low performance, so this is still a challenging issue. In the algorithm perspective, statistical models, such as auto-regressive and regression models, traditional machine learning models, such as SVM and RF, and RNN-based models among the deep learning models have been adopted. To the best of our knowledge, language models, which process the sequential text for language understanding, translation, and generation, have not yet been used in glucose prediction.

III. Method

Our deep learning model framework for glucose prediction is described in this section. We adopted the Transformer model to predict and classify the future blood glucose level of type 2 diabetes patients obtained using a CGM device. The attention mechanism in Transformer will help predict future events such as hyperglycemia and hypoglycemia because a specific value can hint at occurring hyperglycemia and hypoglycemia. We also apply the Transformer to predict the glucose level. To achieve a better performance in the presence of class imbalance naturally occurring in the real world, we propose to use the GAN and generate glucose time series. Furthermore, we will use transfer learning using public data. Type 1 diabetes data will be used because of the large number of samples. The overall framework is depicted in Fig. 1.

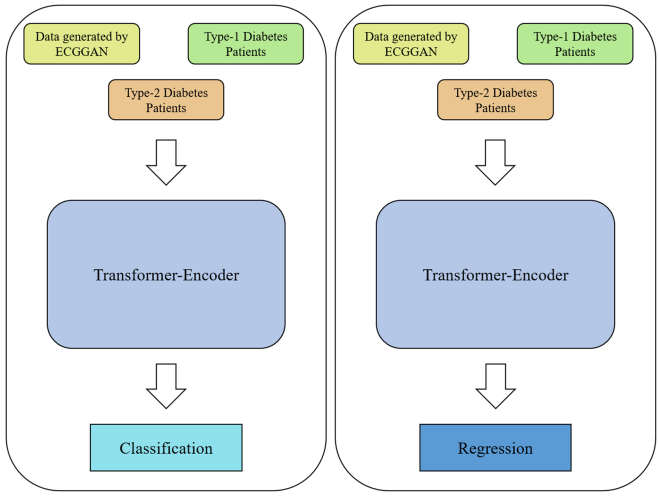

Fig. 1. Research framework.

A. Transformer

The Transformer is a Google-developed model that follows the existing seq2seq encoder - decoder structure but uses attention instead of the RNN structure. This model was created for natural language processing, and it did not employ the RNN or CNN structures in the translation task, instead focusing on essential functions, such as attention and fully connected layers, to achieve good performance. The Transformer is a model relying entirely on self-attention to process the sequential input and transform input to output without using RNN or convolution.

Using the t input and the $t-1$ hidden state to construct the t output, the existing RNN-based model retains the sequential properties of time series data. However, when long-sequence data is used, performance may suffer due to long-term dependency issues. Furthermore, due to the back propagation through time (BPTT), [31] structure of the RNN, there is a difficulty in that sequential operation reduces the operational efficiency. Therefore, the Transformer employs only the attention structure instead of the RNN structure and performs well regardless of the input length. The attention of the Transformer mainly refers to the previous data and finds the critical value to the target (class or value). The output of attention is the weighted sum of input, and this output is connected to the final output layer. This self-attention is expected to be essential in predicting the abnormal class, which is hyperglycemia and hypoglycemia. We aim to develop the prediction model exploiting the attention mechanism in Transformer and determine whether Transformer is effective in classification and regression of glucose of type 2 diabetes. In this study, an experiment was undertaken using the encoder of Transformer as a blood glucose regression and classification model after 30 min, using blood glucose data supplied in 5-min increments. The model utilized in the experiment is explained in this section.

1) Positional Encoding: Unlike the RNN-based model, Transformer has no temporal continuity; hence, additional information about the data placement must be provided to use the sequence of the input data. As a result, material location information is added to the input data via positional encoding. A finite-dimensional representation of a certain point in a sequence is referred to as positional encoding. The outcome of positional encoding for the sequence $A = [a_0, a_1, \ldots a_{n-1}]$ will be a tensor containing information on each element in the sequence as well as the position of elements. Because each component of the input sequence's embedding result and the positional encoding result must be inserted in the original Transformer, the positional encoding result must have the size [sequence length, embedding dimension]. Sinusoidal positional encoding, a structure used in the original Transformer, was utilized in this research. As blood glucose data is a one-dimensional numerical time series, the

input value was turned into a model-dimensional vector without the need for an embedding function, and the positional encoding result was appended and used as an encoder input.

2) Encoder: Only the encoder portion of the Transformer was employed in our experiment. The decoder of the Transformer was not utilized in this study to forecast a single value because it has a structure that constructs each word of the output sentence for the input sentence. Instead, a completely connected layer was added to the end of the encoder to obtain one output value. The number of encoders can be adjusted with num layers, which is the Transformer's hyperparameter, by thinking of the encoder as one layer. Multi-head self-attention and a feed-forward neural network constitute one encoder layer, separated into two sublayers. Self-attention is a method of estimating which features a specific feature refers to within the same context to gain similarity between data in an input sentence. The dot-product is well-implemented and utilized to save space more quickly and efficiently among the attention functions [32]. Scaled dot-product attention prevents the value of Query and Key matrix product from interfering with learning because the weight of the matrix product becomes very large when used. A multi-head calculates multi-head self-attention by executing numerous self-attentions concurrently. This attention allows each head to look at the input from several angles, paying attention to different input sections and capturing more position information. By combining these techniques it is feasible to learn not one but various significant locations. The feed-forward neural network is a fully connected network using the Relu activation function. Before input as the next value, the output value of the multi-head attention layer and the feed-forward neural network layer is subjected to residual connection, and layer normalization [33], [34]. The process of adding the input value of the sublayer's output is called residual connection. Because backpropagation [35] always maintains gradients larger than or equal to 1, this strategy saves the information better. Gradient exploding or vanishing is avoided by layer normalization, and learning is sped up by having a constant value.

B. ECGGAN

Because it is challenging to publish medical data containing personal and sensitive information as public data, the amount of data available is limited, and finding data suited for research is difficult. If there is an imbalance between labels in studies that classify diseases, learning is not carried out correctly, resulting in poor performance. The imbalance can be resolved using existing over-sampling and down-sampling methods, but there will be unavoidable information loss owing to data removal and overfitting issues. Recently, research has been conducted to solve the imbalance problem after augmenting data using generative models. We employed the GAN [36], which is one of several generative models, to supplement insufficient blood glucose data. A generative adversarial network is a generative model that uses competing generators and discriminators to learn the distribution of training data and generate new data. For example, in the GAN learning process for image augmentation, the discriminator learns the actual image to determine the data distribution, and then feeds the latent sample to the generator to genera te a false image. The created fake image is sent into a discriminator trained on original images, and the difference between the two data distributions determines whether it is a fake or real image. When the discriminator receives inputs, the generator develops to make discriminate to judge the fake image erroneously, and the discriminator learns to distinguish the phony image, repeating the process and becoming hostile to each other. To augment blood glucose data with time-series characteristics, we employ ECGGAN [37], which consists of BiLSTM and 1D convolutional layers [38]. BiLSTM improves performance by incorporating reverse learning into the LSTM, reflecting information from the previous historical period. The 1D convolutional layer is utilized because the blood glucose data is one-dimensional time series data. Local features between variables can be extracted in the 1D convolutional layer where the kernel moves with time [39], [40].

IV. Experiments

A. Data Acquisition

We collected data from type 2 diabetic patients using a Dexcom G5(R) Mobile CGM (Decom, Inc., SanDiego, CA, USA) device to evaluate the two models of blood glucose classification and regression. The Dexcom G5 is a smart device with a sensor, transmitter, and mobile app that allows patients to watch glucose readings every 5 min for up to 7 days directly from the continuous real-time sensor (288 samples per day, 2016 samples per week). From April 2019 to January 2022, the target patients were admitted to Soonchunhyang University Cheonan Hospital. All patients met the following requirements. They had diabetes and were aged over 20 but not over 90, including patients in intensive care units. In addition, patients were required to wear a Dexcom G5 for at least 3 to 7 days while in the hospital.

B. Data Statistics

Table II presents the statistics of patients registered for the experiment. The Cheonan Institutional Review Committee approved the construction of the database at Soonchunhyang University Hospital (SCHCA IRB protocol number: SCHCA 2019-11-048). Therefore, the recorded glucose values are stored with a maximum of 400 and a minimum of 40 (mg/dL). The statistics of our data and the public data of type 1 diabetes are included in Table III. Type 2 diabetes has a higher value distribution and a larger standard deviation than type 1, as indicated in Table III. Even though our data come from in-patients whose glucose levels are controlled to be stable by injecting insulin, the variation is higher than that for type 2 diabetes.

C. Data Setting

We conducted an experiment using data from 104 type 2 diabetes patients. The training sample was expanded to include CGM data from 226 type 1 diabetes patients for more accurate model training [41]. Owing to patient movement and technical problems, data may be missing, or outliers may occur among the continuous blood glucose readings captured by the CGM

TABLE I
TAXANOMY OF PREVIOUS RESEARCH ON GLUCOSE PREDICTION USING CGM

Task Type	Reference number/Year	Dataset Type	Number of Samples	Model	Best Performance
Real-Time					
Regression	[16] / 2022	Type 1 Diabetes	12	Rules based learning	MAPE <= 30%
Regression	[17] / 2022	Type 1 Diabetes	12	LR VLSTM BiLSTM Ensemble model	RMSE : 19.63 (mg/dL) (30 min horizon) RMSE : 33.45 (mg/dL) (60 min horizon)
Regression	[19] / 2022	Type 1 Diabetes	12	CRNN SVR Transfer Learning Multi-task Learning	RMSE : 18.8 (mg/dL) MAE : 13.2
Regression	[20] / 2021	Type 1 Diabetes	2	Nonlinear auto regressive neural network	RMSE : 0.695 (mg/dL) (15 min horizon) RMSE : 1.39 (mg/dL) (30 min horizon)
Regression	[21] / 2020	Type 2 Diabetes	20	RNN LSTM GRU	RMSE : 21.5 (mg/dL) MAPE : 11.1%
Regression	[22] / 2020	Type 1 Diabetes	49	LSTM	RMSE : 6.42 (mg/dL) (30 min horizon) RMSE : 11.35 (mg/dL) (60 min horizon) RMSE : 3.45 (mg/dL) (simulated patient, 30 min horizon)
Regression	[23] / 2020	Type 1 Diabetes	16	DRNN	RMSE : 18.9 (mg/dL)
Regression	[25] / 2020	Type 1 Diabetes	20	CRNN	RMSE : 21.07 (mg/dL) (30 min horizon) MARD: 11.61% (30 min horizon) RMSE : 33.27 (mg/dL) (60 min horizon) MARD : 19.01% (60 min horizon)
Classification	[26] / 2021	Type 1 Diabetes Type 2 Diabetes	12 40	RNN CNN SAN Transfer Learning	AUC : 95.98% Sensitivity : 59.19% Specificity : 98.15% F1-Score : 0.61
Classification	[27] / 2019	Type 1 Diabetes Type 2 Diabetes	52 52	RF SVM KNN LR	AUC : 96.6% Sensitivity : 89.6% Specificity : 91.3% F1-Score : 0.54
Non Real-Time					
Classification	[28] / 2020	Type 1 Diabetes	10	MLP SVM	AUC : 80% Sensitivity : 78% Specificity : 82%
Classification	[29] / 2020	Type 1 Diabetes	463	LDA	Sensitivity : 75% Specificity : 70%
Classification	[30] / 2017	Type 1 Diabetes	11	KNN	Hypoglycemia AUC : 83.09% Sensitivity : 88.78% Hyperglycemia AUC : 83.64% Sensitivity : 82.58%

TABLE II
TYPE 1 AND TYPE 2 DIABETES PATIENT DEMOGRAPHICS

Dataset Type	Age	Sex	Count	Total	Glucose Length Average
Type 2 Diabetes	20 - 29	Female	1	40	1393.2
	30 - 39		9		
	40 - 49		7		
	50 - 59		12		
	60 - 69		9		
	70 - 79		2		
	20 - 29	Male	3	64	
	30 - 39		5		
	40 - 49		10		
	50 - 59		17		
	60 - 69		20		
	70 - 79		9		
Type 1 Diabetes	44 ± 14	Female	112		66153.3
		Male	114		

TABLE III
TYPE 1 AND TYPE 2 GLUCOSE DESCRIPTIVE STATISTICS

Data Type	Number Of Samples	Min	Q1	Median	Mean	Q3	Max	Std
Type 1 Diabetes	226	40	112	150	160.3	198	400	64.67
Type 2 Diabetes	104	40	131	172	183.3	224	400	68.95

device. If the blood glucose value was not numerical, it was deleted from the equation, and the maximum and minimum values were set to 400 (mg/dL) and 40 (mg/dL), respectively. The out-of-range values were translated to 400 (mg/dL) and 40 (mg/dL), respectively. The data were reconstructed as a supervised learning task because it is important to predict and classify future blood glucose values based on the historical levels. Because blood glucose levels are time-series data stored over time, the data input component of a model should be set using a sliding window approach, as shown in Fig. 4.

The input length was set to 12 steps (60 min), 24 steps (120 min), and 36 steps (180 min) in this study, whereas the model's output value was set to a single blood glucose level or class based on the glucose level after six steps (30 min). We also performed the experiment with three (15 min) and nine (45 min) steps to compare the performance differences for delay time. The blood glucose regression was evaluated by comparing the model's predicted value to the actual blood glucose level through the MAPE. For classification, if the blood glucose value after 30 min was higher 180, the class was defined as hyperglycemia; if it was lower than 70, the class was hypoglycemia; and all other values were classed as normoglycemia. In the preprocessing process of labeling the data into three classes, it was noticed that the ratios of hyperglycemia, hypoglycemia, and normoglycemia were not equally distributed, as indicated in Table IV.

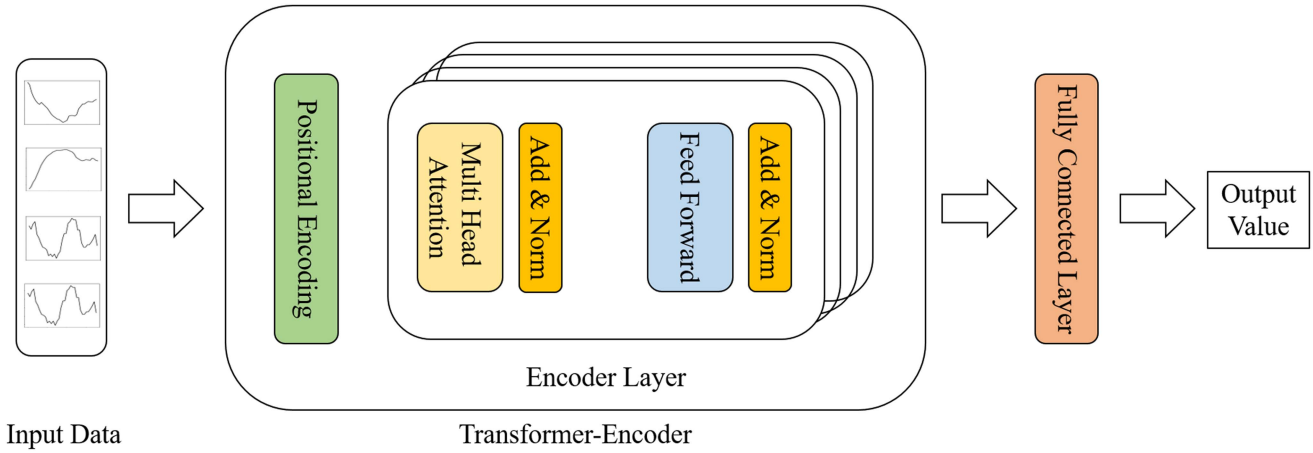

Fig. 2. Transformer Encoder model structure for classification and regression of blood glucose levels. After passing through the fully connected layer, a predicted blood glucose value is calculated after 30 min, and in the case of blood glucose categorization, it is estimated as one of normoglycemia, hypoglycemia, or hyperglycemia after 30 min.

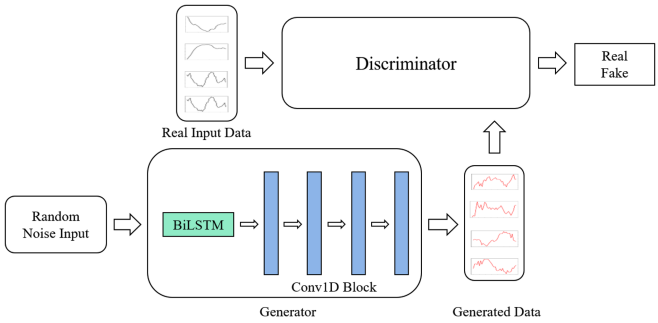

Fig. 3. ECGGAN structure used for data augmentation.

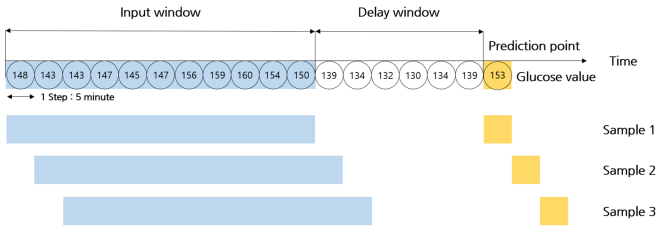

Fig. 4. Input data generation process.

TABLE IV
RATIO OF EACH CLASS WITH INPUT LENGTH OF 36 STEPS

Class	Ratio
Normoglycemia	56.87%
Hyperglycemia	42.26%
Hypoglycemia	0.89%

Because the data were acquired from actual hospitalized patients, the hypoglycemia class is very modest compared to the other two classifications. Critical problems such as hypoglycemic shock might arise if hypoglycemia occurs while

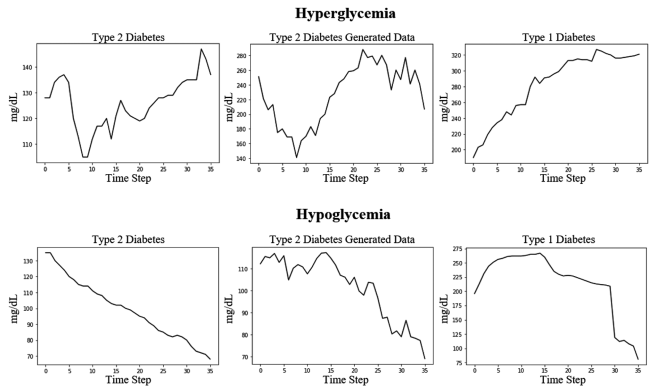

Fig. 5. Visualization of a blood glucose sample using a 36 input length.

collecting data from patients in clinical practice. When hypoglycemia occurs, the medical personnel responds quickly and tries to normalize the glucose level, leading to a low proportion of hypoglycemia class. When there is a class imbalance like this, achieving good performance in the classification task is difficult. Therefore, the hypoglycemia and hyperglycemia data, which are insufficient compared to normal glucose levels, were enhanced to match the normoglycemia class using the ECGGAN. Some hypoglycemia and hyperglycemia data among the entire classes are visualized and shown in Fig. 6. The graph on the left shows blood glucose data from a type 2 diabetes patient, while the middle graph shows data obtained using ECGGAN to augment type 2 diabetes data. The type 1 diabetes data in the last graph on the right is additional data used to train the model. The generated data have a less smooth time-series, so we expected that the generated data would not be helpful for regression. We confirmed our expectations through the experiment. However, the generated data mimicked the hyperglycemia and hypoglycemia patterns well, like the original data, and the patterns from generated data were more similar to those of the real data than

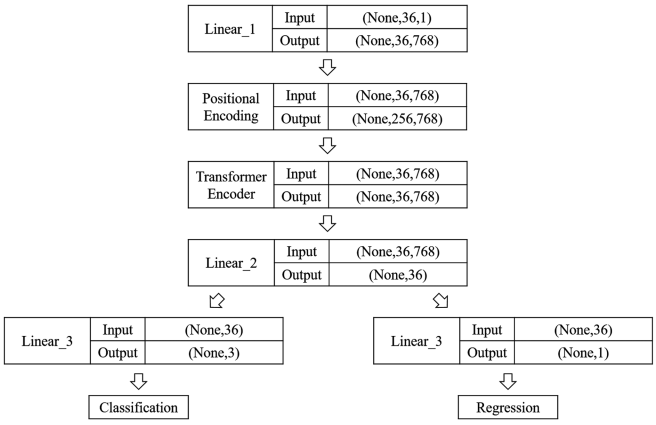

Fig. 6. Model architecture of input length of 36.

TABLE V
MMD METRICS TO ASSESS THE PERFORMANCE OF AUGMENTED DATA

Metric	Class	12 Step	24 Step	36 Step
MMD	Hypoglycemia	3.42×10^{-4}	2.63×10^{-4}	5.42×10^{-4}
	Hyperglycemia	3.42×10^{-3}	8.33×10^{-3}	2.43×10^{-3}

the type 1 data. This will be helpful for the hyperglycemia and hypoglycemia classification.

D. Result

We evaluated the performance outcomes for the Transformer and ECGGAN models. The maximum mean discrepancy (MMD) [42] evaluation score was used to assess the data produced by ECGGAN. By mapping the source and target in reproducing the kernel Hilbert space and then determining the distribution difference, the MMD evaluation score is a formula that indicates the inconsistency of the two distributions. Table V lists the average of the MMD of the generated data and actual data for hypoglycemia and hyperglycemia based on the input length. We found that the distributions of the generated data and actual samples were sufficiently similar, but the MMD score varied according to the input length and type of classes. The ECGGAN model for hypoglycemia enhancement was tested with a generator consisting of one BiLSTM layer with a hidden size of 128 and five Conv1D layers and a discriminator consisting of four Conv1D layers and one output layer when the input length was 24 steps.

The experiment was performed by dividing the training and test data in a 7:3 ratio. We changed the parameters to discover the best performance. For both tasks, the epoch was set to 100, and early stopping was used to prevent overfitting of the training data by halting the training when the loss value did not fall. We set the batch size to [64, 128, 256] and run an experiment on each batch size to see how well it performed. The scheduler was implemented by setting the learning rate to [1e-03, 1e-04, 1e-05], and the learning rate was changed at scheduled epochs. The Adam optimizer was used as an optimization algorithm. To find out the best architecture of the Transformer for glucose regression and classification, hyper-parameter tuning was performed as follows. Experiments were carried out to set the number of encoder layers in the model to [1, 2, 3, 4], and the encoder's input dimension model to [128, 256, 512, 768], and as well as the number of heads to be used in parallel in multi-head attention was set to [1, 2, 4, 8].

TABLE VI
EXPERIMENT RESULTS FROM BLOOD GLUCOSE REGRESSION MODEL

Metrics	Delay Time	Input Step	Type 2 Diabetes	Type 2 Diabetes + Type 1 Diabetes
MAPE	15	12	19.46	10.24
		24	11.58	10.58
		36	10.3	10.0
	30	12	17.89	12.78
		24	16.82	13.4
		36	16.4	13.5
	45	12	31.44	24.11
		24	23.35	22.18
		36	19.87	23.61

The Transformer model comprised 768 model input dimensions, four multi-heads, and it was one layer when the input length was 36 steps, and it had relatively high performance in the classification experiment. However, when the input length was 12 steps, the regression experiment consisted of 512 model input dimensions, four multi-heads, and one layer. The architecture of the proposed model is displayed in Fig. 6.

1) Blood Glucose Regression: In this experiment, we used the MAPE evaluation index to compare the difference between the actual blood glucose level and the projected value for each input part. In addition, the prediction model's performance using only the type 2 diabetes patient data we collected and the performance of the model additionally utilizing type 1 diabetes data were assessed and compared. The model's performance with additional data on type 1 diabetes was better than the model with only type 2 diabetes during the model training procedure, as indicated in Table VI. As observed, the forecasting performance for 15 min advance is the best. As the delay windows become longer, the performance drops. The length of the input step shows the contradicting performance. For the longer delay time, the longer input step shows the best performance. However, the shorter input step is appropriate in forecasting with 30 min advance, which is our goal.

2) Blood Glucose Classification: We used precision and recall to evaluate the performance in the classiffication of normoglycemia, hypoglycemia, and hyperglycemia. Because there are fewer hypoglycemia samples, the high accuracy does not guarantee that hypoglycemia is genuinely discriminated. The experiment is composed of three parts based on the input length. The performance comparison was made using only type 2 diabetes patient data, data augmentation with the ECGGAN model to resolve unbalanced data, and data augmentation and extra training with type 1 diabetes data. The results are presented in Table VII. The performance for hypoglycemia is very bad when using only the type 2 dataset. When the number of samples is increased by incorporating the type 1 dataset, the performance increases overall and dramatically for hypoglycemia. For classification, the input step of 36 for 15 min ahead prediction generated the best performance. For 30 min ahead prediction, the input length of 24 was the best. We concluded that the reference

TABLE VII
EXPERIMENT RESULTS FROM BLOOD GLUCOSE CLASSIFICATION MODEL

Metric	Delay Time	Input Step	Type 2 Diabetes			Type 2 Diabetes + Generated Data			Type 2 Diabetes + Generated Data + Type 1 Diabetes		
			Normoglycemia	Hyperglycemia	Hypoglycemia	Normoglycemia	Hyperglycemia	Hypoglycemia	Normoglycemia	Hyperglycemia	Hypoglycemia
Precision	15	12	0.56	0.72	0.0	0.94	0.92	0.59	0.89	0.91	0.58
Recall			0.96	0.12	0.0	0.93	0.93	0.65	0.92	0.88	0.49
Precision		24	0.91	0.5	0.52	0.92	0.95	0.43	0.94	0.9	0.73
Recall			0.18	1.0	0.03	0.95	0.90	0.81	0.94	0.94	0.49
Precision		36	0.64	0.65	0.0	0.94	0.92	0.61	0.94	0.93	0.66
Recall			0.79	0.48	0.0	0.93	0.93	0.43	0.94	0.94	0.68
Precision	30	12	0.58	0.67	0.0	0.85	0.93	0.40	0.89	0.91	0.58
Recall			0.67	0.20	0.0	0.95	0.82	0.20	0.92	0.88	0.45
Precision		24	0.86	0.91	0.0	0.88	0.91	0.37	0.89	0.92	0.57
Recall			0.91	0.83	0.0	0.93	0.86	0.05	0.93	0.87	0.45
Precision		36	0.67	0.73	0.0	0.88	0.77	0.38	0.89	0.93	0.59
Recall			0.84	0.51	0.0	0.77	0.88	0.36	0.93	0.87	0.47
Precision	45	12	0.58	0.67	0.0	0.86	0.86	0.40	0.84	0.90	0.56
Recall			0.91	0.22	0.0	0.88	0.85	0.21	0.92	0.81	0.04
Precision		24	0.80	0.54	0.0	0.84	0.89	0.23	0.83	0.91	0.51
Recall			0.36	0.91	0.0	0.91	0.80	0.23	0.94	0.79	0.05
Precision		36	0.62	0.69	0.0	0.82	0.60	0.0	0.85	0.90	0.59
Recall			0.86	0.38	0.0	0.53	0.88	0.0	0.93	0.82	0.16

to a very distant past rather decreases the performance in the experiments.

V. DISCUSSION

The additional type 1 diabetes data used to train the model were collected from adults with type 1 diabetes using a Dexcom G4 device between May 2015 and March 2016. The participants were 44 years old on average, and data were collected from 226 people. Blood glucose levels were recorded at 5-min intervals from 40 (mg/dL) to 400 (mg/dL) for up to 7 days.

There are several variables in addition to the CGM variable, such as blood glucose monitoring (BGM) and glycated hemoglobin (HbA1C), but we only used the blood glucose level measured by the CGM device. When the input length was 12 steps, and type 1 diabetes data was additionally included, the average MAPE of 12.78% performed the best in the regression experiment. However, if we only used the type 2 diabetes data, the performance improved without degrading even for a considerable input length. In addition, the number of samples increased when type 1 diabetes data were included during model training, allowing for the training of additional samples, which improved performance. Finally, individual patient data were fed into the trained model and visualized to compare the best and worst performances. The visualization results are shown in Fig. 7.

The worst performing result's MAPE was 28.97%, while the best performing result's MAPE was 8.15%. A comparison of the two figures shows that the performances for hypoglycemia and irregular blood glucose variations are significantly reduced. In the baseline experiment, where the last value of the input length was compared with the value after 30 min without the training model, a 15.43% MAPE was attained when the input length was 12. In comparison to the baseline result, the performance of our experiment is better.

We used Clarke error grid to analyze the results that showed the best performance in Fig. 8. With the best-performance model, we carried out error analysis by comparing the actual and predicted blood glucose levels. The Clarke grid divides the scatter plot of the actual and predicted blood glucose levels into five sections. The error analysis results show that most errors are concentrated in region A and are only moderately scattered in area B. Area A is the region where the difference between

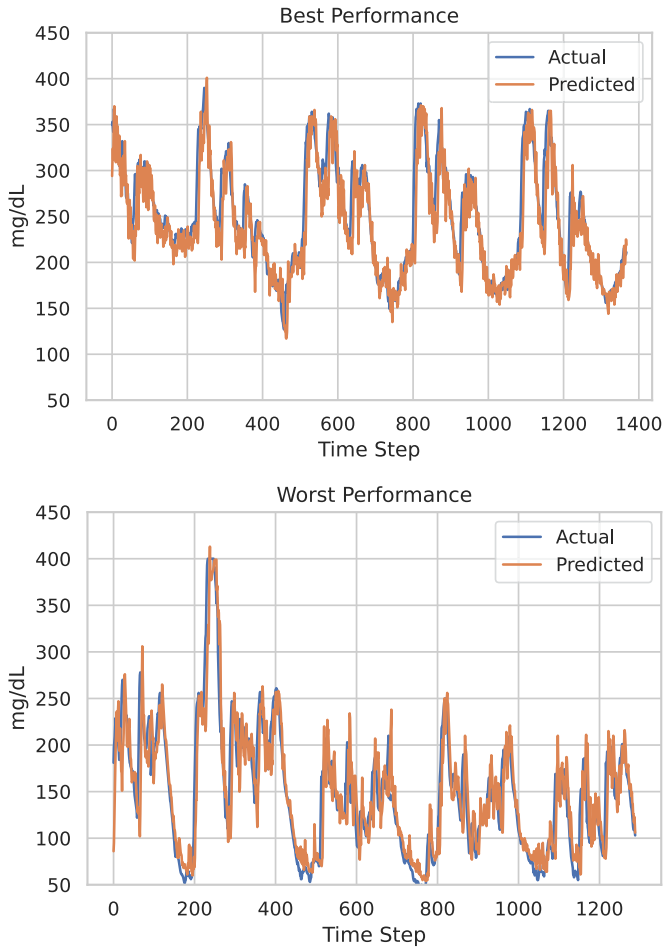

Fig. 7. Visualization of actual and predicted values. The figure on the top shows the best performance, and the figure on the bottom shows the worst performance.

the predicted and actual blood glucose levels is below 20%. The discrepancy between the two readings in area B is higher than 20% but still falls within the standard limits. In area D, where the forecast is inaccurate, a relatively small number of values are distributed. Because area B contains many points, if the actual value is high, the predicted value appears lower than it in some regions. We also evaluated the test performance on the additional

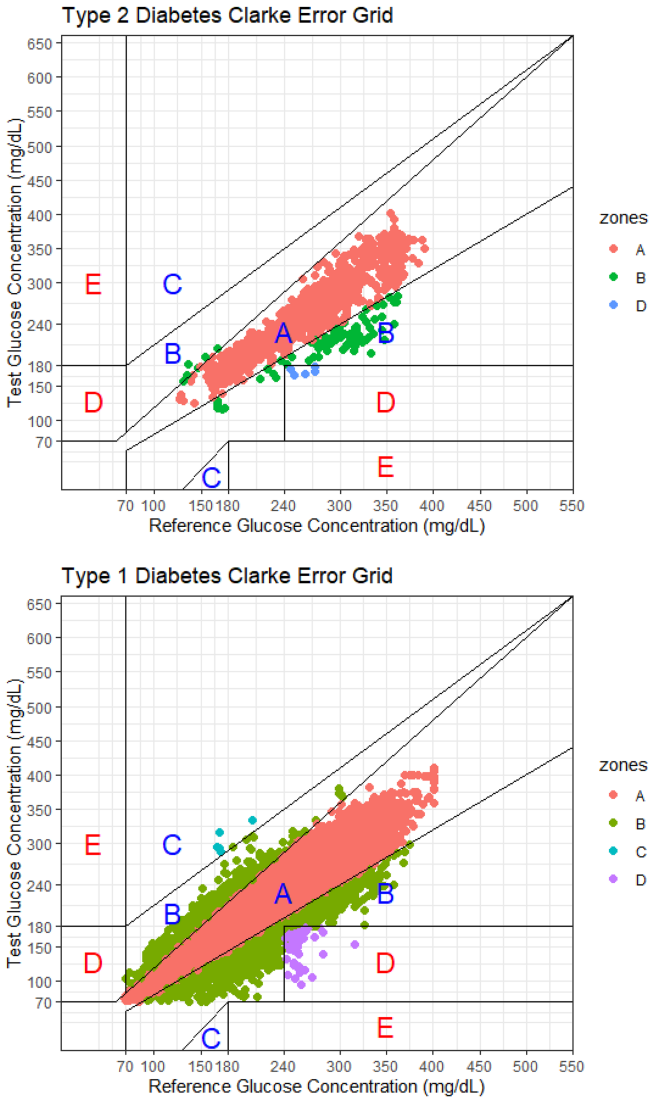

Fig. 8. Error analysis using the actual blood glucose levels (x - axis) and the predicted blood glucose levels (y - axis).

TABLE VIII
DISTRIBUTION RATIO OF GRID AREA

Dataset Type	Zone	A	B	C	D	E
Type 2 Diabetes	Clarke	0.9386	0.0577	0	0.0037	0
Type 1 Diabetes		0.9617	0.0372	0.0001	0.0011	0

type 1 diabetes data used for training. The average MAPE of type 1 diabetes was 29.84%, and the best performance value was 7.66% MAPE. The Clarke error grid analysis was performed at the value showing the best performance. Table VIII confirmed that most values corresponded to A, and the rest were distributed in the order of B, D, and C.

To evaluate the reliability of the model, we conducted 5-fold cross-validation using only our type 2 diabetes data to evaluate the generalized performance. We split the dataset into 5-fold, built the model five times, and tested the models. Regression

TABLE IX
5-FOLD CROSS VALIDATION RESULTS

Cross Validation	Model	Experiment	Metrics	12 Step	24 Step	36 Step
5-Fold	Transformer	Regression	MAPE	13.76	15.36	14.94
			STD	0.32	0.7	0.85
		Classification	F1 Score	0.68	0.66	0.71
			STD	0.02	0.036	0.01

TABLE X
RESULTS OF TRANSFORMER AND GRU WITH TYPE 1 DIABETES DATA AND AUGMENTED DATA IN TERMS OF F1 SCORE FOR CLASSIFICATION AND MAPE FOR REGRESSION

Experiment	Model	12 Step	24 Step	36 Step
Regression	Transformer(our)	12.78	13.4	13.5
	GRU	15.60	15.49	15.57
Classification	Transformer(our)	0.77	0.76	0.77
	GRU	0.72	0.73	0.69

and classification were evaluated by MAPE and F1 score, respectively, and standard deviation (STD) indicators were also utilized to verify the performance deviation for each fold. The experiment was conducted at a delay time of 30 min, and the results are represented in Table IX. The experimental results showed better performance than the fixed test dataset. STD for regression and classification is low. We concluded that the results presented in other tables without cross-validation are reliable.

When type 1 diabetes data and augmented data were used in the Transformer model for the classification experiment, the results in all experiments exhibited better performance than in those using type 2 diabetes only. It is apparent that, before data augmentation, it was challenging to identify the hypoglycemic class because its ratio was much smaller than those of the normoglycemic and hyperglycemia classes. When type 1 diabetes was included in the data during model training, the model could learn the pattern of the hypoglycemia class and identify the hypoglycemic class when assessing the test performance. Furthermore, the data used in training increased and showed better experimental results than that in the regression experiment.

Using the same dataset, we compared our model with other models, including time-sequence-based deep learning. The GRU, our reference model, is an RNN-based model, a slightly simplified version of the LSTM's time-step cell. It trains faster than LSTM and performs similarly.

The experiment results presented in Table X showed that the Transformer outperformed the GRU in the 30 min delay setting. Our model could not provide the SOTA for regression, but we achieved a better performance in classification. Therefore, we believe that Transformer is more effective for classification than regression in the case of time-series data. We demonstrated excellent performance in the experiment by only using the blood glucose level as an input. We also applied the public dataset, OhioT1DM [43], to the Transformer trained through the transfer learning to evaluate its performance. The OhioT1DM dataset was collected for eight weeks from 12 patients with type 1 diabetes and is the first open blood glucose data set that includes continuous glucose monitoring data, meal information and sleep time. When we tested our model using the OhioT1DM dataset, the evaluation results were better than those of previous

TABLE XI
MODEL EVALUATION USING OHIOT1DM DATASET

Dataset	Model	Experiments	12 Step	24 Step	36 Step
OhioT1DM	Transformer	Classification	0.72	0.70	0.75
		Regression	17.88	18.25	18.47

TABLE XII
CLASSIFICATION AND REGRESSION MODEL COMPARISON EXPERIMENT FOR TYPE2-ONLY

Model	MAPE	F1 Score
Transformer(our)	17.89	0.67
Random Forest	22.73	0.60
Support Vector Machine	21.71	0.61

works [26] and [27] in the classification task. The results are shown in Table XI.

In addition, we also compared the performance of the regression model with the machine learning model. RF is an ensemble technique that learns multiple decision trees, and SVM is a classifier that finds the decision boundary as far away as possible from the two classes and is a method of classifying classes while satisfying certain conditions. For a quick comparison, this comparative experiment used only type 2 diabetes data when the delay time was 30 min and the input length was 12, which was the setting that generated the worst performance. The experimental results indicated that our model performs better when compared to other machine learning models. For classification, we also adopted the SVM and RF. They are known as good-performing algorithms, the latter belongs to the assemble algorithms whereas the former does not. We tested other models for the worst-performed setting with type 2, but we found that our model outperformed other traditional models. The experiment results are demonstrated in Table XII.

Our model has drawbacks. First, the Transformer is heavier than other models referenced in this paper, and thus it requires much more time for training. When updating the previously built model, it is not recommended to update the model frequently, while other models can be updated whenever new data is collected. Second, our model was built using CGM only, even though other time-sequence data are available. Although it was not used in the experiment, the collected data included whether or not the patient had eaten or had been administered insulin. This was not utilized in the experiment because, although the patients' mealtimes are set, the amount and rate of their intake vary from patient to patient, and it is challenging to record food consumption outside of mealtimes. In future research, we intend to collect in a precise way the variables that affect blood glucose levels, such as whether or not food has been eaten or insulin injected, as we believe that additional information beyond blood glucose levels will help predict these levels. Third, we just tested one of GAN models for time-sequence data, even tough several models have been proposed for time-series data augmentation. Because the data were collected by observing hospitalized patients for 3 to 7 days, there is a limit on how long they may feed data into the model in our experiment. We intend to increase the data collection period to identify performance degradation with increased input length. The ratio of the hypoglycemic class among the data gathered is relatively low, as could be observed when the blood glucose level was labeled for the classification experiment. Because the data were taken from actual inpatients, it was impossible to raise the hypoglycemia class purposefully. To address this imbalance in our experiment, we used the ECGGAN model, but we also intend to assess the performance using various augmented models, such as TimeGAN [44].

VI. CONCLUSION

In this study, we developed a blood glucose prediction model for patients who have type 2 diabetes using the language model, Transformer. Our model intends to help physicians reduce their labor force while also allowing them to prepare appropriately for the occurrence of dangerous situations. Most blood glucose research has been conducted on patients with type 1 diabetes. However, it is still essential to monitor and control the glucose level of type 2 diabetes patients, primarily inpatients. The regression model and classification model were developed and tested to predict the blood glucose level after 30 min based on the blood glucose data from the CGM device. Unlike prior experiments with RNN-based models, in which performance degrades as the input length grows longer, performance loss due to input length did not occur significantly in our Transformer experiment. Because data were obtained from actual patients, there is a data imbalance, but the problem was overcome using the ECGGAN model. Furthermore, type 1 diabetes data was employed as transfer learning, preventing performance loss due to less training data when the model was trained using only type 2 diabetes data.

REFERENCES

[1] American Diabetes Association, "Diagnosis and classification of diabetes mellitus," *Diabetes Care*, vol. 32, no. Suppl. 1, pp. S62–S67, Jan. 2009.
[2] International Diabetes Federation, *IDF Diabetes Atlas*, 10th ed. Brussels, Belgium: Int. Diabetes Federation, 2021.
[3] American Diabetes Association Professional Practice Committee, "6 Glycemic Targets: Standards of Medical Care in Diabetes—2022," *Diabetes Care*, vol. 45, no. Suppl_1, pp. S83–S96, Jan. 2022.
[4] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.
[5] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2021.
[6] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci.*, vol. 79, no. 8, pp. 2554–2558, 1982.
[7] M. Martin et al., "Irinagain/Awesome-CGM: List of public CGM datasets (version v1.1.0)," *Zenodo*, Apr. 27, 2021, doi: 10.5281/zenodo.4723654.
[8] A. Earnest, M. I. Chen, D. Ng, and L. Y. Sin, "Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore," *BMC Health Serv. Res.*, vol. 5, no. 1, May 2005, Art. no. 36.
[9] J. Saejong and C. Hyungjun, "Seasonal time series forecasting using machine learning," *J. Korean Data Anal. Soc.*, vol. 22, no. 5, pp. 1779–1791, 2020.
[10] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, 2015.
[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[13] J. Chung, Ç. Gülçehre, K. H. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. Neural Inf. Process. Syst. Workshop Deep Learn.*, Dec. 2014, vol. 2014.
[14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, Cambridge, MA, USA, 2014, vol. 2, pp. 3104–3112.
[15] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. 3rd Int. Conf. Learn. Representations*, Y. Bengio and Y. LeCun, Eds. San Diego, CA, USA, May 7-9, 2015.
[16] T. Koutny and M. Mayo, "Predicting glucose level with an adapted branch predictor," *Comput. Biol. Med.*, vol. 145, Jun. 2022, Art. no. 105388.
[17] H. Nemat, H. Khadem, M. R. Eissa, J. Elliott, and M. Benaissa, "Blood glucose level prediction: Advanced deep-ensemble learning approach," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 6, pp. 2758–2769, Jun. 2022.
[18] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2005, vol. 4, pp. 2047–2052.
[19] J. Daniels, P. Herrero, and P. Georgiou, "A multitask learning approach to personalized blood glucose prediction," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 1, pp. 436–445, Jan. 2022.
[20] M. Asad, U. Qamar, and M. Abbas, "Blood glucose level prediction of diabetic type 1 patients using nonlinear autoregressive neural networks," *J. Healthcare Eng.*, vol. 2021, Feb. 2021, Art. no. 6611091.
[21] D.-Y. Kim et al., "Developing an individual glucose prediction model using recurrent neural network," *Sensors*, vol. 20, no. 22, Nov. 2020, Art. no. E6460.
[22] M. Munoz-Organero, "Deep physiological model for blood glucose prediction in T1DM patients," *Sensors*, vol. 20, no. 14, Jul. 2020, Art. no. 3896.
[23] T. Zhu, K. Li, J. Chen, P. Herrero, and P. Georgiou, "Dilated recurrent neural networks for glucose forecasting in type 1 diabetes," *J. Healthcare Inform. Res.*, vol. 4, pp. 308–324, 2020.
[24] S. Chang et al., "Dilated recurrent neural networks," in *Proc. Neural Inf. Process. Syst. Found.*, 2017, pp. 76–86.
[25] K. Li, J. Daniels, C. Liu, P. Herrero, and P. Georgiou, "Convolutional recurrent neural networks for glucose prediction," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 2, pp. 603–613, Feb. 2020.
[26] Y. Deng et al., "Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients," *NPJ Digit. Med.*, vol. 4, no. 1, pp. 1–13, Jul. 2021.
[27] W. Seo, Y.-B. Lee, S. Lee, S.-M. Jin, and S.-M. Park, "A machine-learning approach to predict postprandial hypoglycemia," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, Nov. 2019, Art. no. 210.
[28] A. Bertachi et al., "Prediction of nocturnal hypoglycemia in adults with type 1 diabetes under multiple daily injections using continuous glucose monitoring and physical activity monitor," *Sensors*, vol. 20, no. 6, Jan. 2020, Art. no. 1705.
[29] M. H. Jensen, C. Dethlefsen, P. Vestergaard, and O. Hejlesen, "Prediction of nocturnal hypoglycemia from continuous glucose monitoring data in people with type 1 diabetes: A proof-of-concept study," *J. Diabetes Sci. Technol.*, vol. 14, no. 2, pp. 250–256, Mar. 2020.
[30] N. Mordvanyuk, F. Torrent-Fontbona, and B. Lopez, "Prediction of glucose level conditions from sequential data," in *Recent Advances in Artificial Intelligence Research and Development*, IOS Press, 2017, pp. 227–232.
[31] P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.
[32] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Lisbon, Portugal, Sep. 2015, pp. 1412–1421.
[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Los Alamitos, CA, USA, Jun. 2016, pp. 770–778.
[34] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *CoRR*, vol. abs/1607.06450, 2016.
[35] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
[36] I. J. Goodfellow et al., "Generative adversarial networks," in *Advances in Neural Information Processing Systems*, vol. 24, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014.
[37] A. M. Delaney, E. Brophy, and T. E. Ward, "Synthesis of realistic ECG using generative adversarial networks," 2019.
[38] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, 2021, Art. no. 107398.
[39] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional gans," Dec. 4, 2017.
[40] F. Zhu, F. Ye, Y. Fu, Q. Liu, and B. Shen, "Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network," *Sci. Rep.*, vol. 9, no. 1, May 2019, Art. no. 6734.
[41] G. Aleppo et al., "REPLACE-BG: A randomized trial comparing continuous glucose monitoring with and without routine blood glucose monitoring in adults with well-controlled type 1 diabetes," *Diabetes Care*, vol. 40, no. 4, pp. 538–545, Feb. 2017.
[42] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, no. 25, pp. 723–773, 2012.
[43] C. Marling and R. Bunescu, "The OhioT1DM dataset for blood glucose level prediction: Update 2020," in *Proc. CEUR Workshop*, 2020, pp. 71–74.
[44] J. Yoon, D. Jarrett, and M. V. D. Schaar, "Time-series generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, vol. 32.