

MTC-GA: A multitasking approach for blood glucose level forecasting.

Mennatullah Mahmoud, Walid Al-Atabany and Ahmed Soltan *Senior, IEEE*

Abstract—Effective diabetes management depends significantly on precise glucose monitoring to prevent critical incidents, with Continuous Glucose Monitoring (CGM) systems being essential. Deep learning models for blood glucose prediction can improve CGM by facilitating proactive actions; nevertheless, existing methodologies depend on either highly accurate but complex personalized models or less precise stacked/generalized models. This study presents MTC-GA, a hybrid approach of employing a multitasking model while being integrated with Temporal Convolutional Networks (TCN), Gated Recurrent Units (GRU), and multi-head attention within the task-specific layers. This methodology achieves an accuracy similar to personalized models while decreasing complexity by almost 40% making it more suitable in real-world applications. The model is analyzed in comparison to personalized and generalized models over prediction horizons (PHs) ranging from 15 to 120 minutes, increasing in 15-minute periods. Besides this, it is clinically evaluated using Clarke Grid Error (CEG) Analysis along with other statistical methods. It achieved the lowest numerical and clinical errors compared to the presented models in the literature. This demonstrates a Mean Absolute Error (MAE) between 6.77 and 33.70 and a Root Mean Square Error (RMSE) ranging from 10.36 to 43.46, showing its capacity for scalable and effective diabetes treatment.

Index Terms—Time series forecasting, Glucose forecasting, Multitasking model, Multihead attention, Gated recurrent unit, Temporal convolutional network.

I. INTRODUCTION

HEALTH monitoring solutions based on the Internet of Things (IoT) were recently established to improve the quality of health care services [1]. These technologies have transformed preventative healthcare by facilitating early diagnosis of medical disorders [2], ongoing monitoring of chronic diseases [3], and individualized treatment options [4], among other advancements. Continuous Glucose Monitoring (CGM) is a critical tool for diabetes management and one of the most transformational IoT applications in health monitoring. It commonly employs minimally invasive or non-invasive sensors to assess glucose levels in real-time offering detailed readings indicating patient's glycemic patterns [5]. By delivering continuous data, CGM enables patients and healthcare providers to optimize insulin dosage, mitigate hypoglycemia

This work is supported by the Information Technology Academia Collaboration (ITAC) program under grant number CFP225. (Corresponding author: Mennatullah Mahmoud.)

Mennatullah Mahmoud and Ahmed Soltan are with Nanoelectronics Integrated Systems Center (NISC), Nile University, 26th of July Corridor, Sheikh Zayed City, 12588, Giza, Egypt (e-mail: me.amin@nu.edu.eg; asoltan@nu.edu.eg).

Walid Al-Atabany is with Biomedical informatics program, School of Information Technology and Computer Science (ITCS), Nile University, Sheikh Zayed City, 12588, Giza, Egypt

or hyperglycemia risk and improve overall glycemic regulation [6]. Nonetheless, the substantial volume and intricacy of CGM data provide difficulties for both patients and healthcare practitioners in deriving relevant insights.

Artificial intelligence (AI) has emerged as a significant improvement facilitating enhanced data interpretation and early detection of critical glucose events. Forecasting blood glucose levels enables early detection of hypoglycemia and hyperglycemia which present acute hazards to patients [7]. Hypoglycemia is characterized by critically low blood glucose levels often below the normal threshold of 70 mg/dL (3.9 mmol/L). It is deemed a medical emergency if untreated, as it may result in significant complications in the short and long term. This includes loss of consciousness, seizures, neurological damage from repeated severe episodes or leading to a condition called "Hypoglycemia unawareness" where symptoms are not felt, increasing danger [8]. Conversely, hyperglycemia, characterized by high blood glucose levels of >140 mg/dL (7.8 mmol/L) fasting or >180 mg/dL (10 mmol/L) postprandial. Chronic hyperglycemia is a hallmark of diabetes where its complications can be very severe. For short-term risks, it may lead to Diabetic Ketoacidosis (DKA) which is life-threatening condition in type 1 diabetes characterized by acidosis and dehydration [9]. For type 2 diabetes, it leads to Hyperosmolar Hyperglycemic State (HHS) with severe dehydration leading to confusion, seizures, or coma [9]. For the most dangerous part is the long-term risks where chronic complications can happen as microvascular damage including Retinopathy (blindness), nephropathy (kidney failure), and neuropathy (nerve pain or dysfunction) or even macrovascular damage as cardiovascular diseases, including heart attacks and strokes. Moreover, other risks may happen such as weakened immune response, amputation risk or cognitive decline as dementia and Alzheimer's disease [10].

By providing real-time alerts through connected devices, AI-powered CGM systems empower patients to enable early intervention, allowing users to take preventive actions before critical glucose levels are reached. For prediction algorithms, there are three primary categories: physiological models, hybrid models and data-driven models (black-box methodologies). Data-driven models are attracting significant interest and are being increasingly investigated. These models can be categorized into classical time series forecasting (CTF), traditional machine learning (TML), and deep neural network (DNN) methodologies [11]. In the context of BGL forecasting, various research has been done using DNN methodologies for being the most effective approach [12]. Exploring these methodologies, it is found that BGL forecasting papers can be categorized

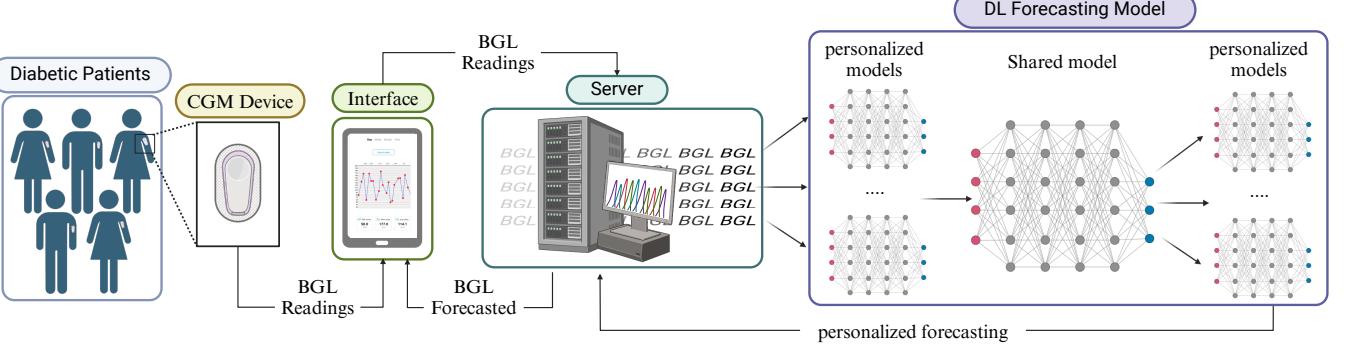


Fig. 1. The data flow from diabetic patients wearing CGM devices to personalized forecasting outputs via server-based processing and multitasking model merging between both personalized and shared models.

by employing Feedforward Forward Network (FNN) [13], [14] and Convolutional Neural Network (CNN) [15], [16] either as standalone model or embedded layers. While few studies employed temporal convolutional network (TCN) [17], [18], conversely to recurrent neural network (RNN) which is the most widely used algorithm in BL forecasting with all its variants [17], [19]. Other studies use a self-attention netwrok (SAN) in their model which is often associated with the transformer architecture [20]–[22]. Despite all these advancements, several gaps remain in the current literature that limit their wider applicability and reliability. A significant issue is the lack of testing of models over diverse prediction horizons (PHs) resulting in being unclear regarding their robustness in different temporal contexts. Moreover, several studies depend on limited evaluation metrics failing to provide a comprehensive assessment of the model performance. Few studies have evaluated their model performance compared to state-of-the-art methodologies. Moreover, certain research employs supplementary features including carbohydrate intake and insulin dosages, into their models. Although these features can improve predictive accuracy, their availability is frequently constrained in real-world situations making such models impracticable for general use. Ultimately, the majority of studies follow a training methodology of either personalized or generalized (stacked) which they both lack reliability in practical scenarios. Addressing these gaps is essential to ensure the development of AI models that are robust, practical, and effective in diverse real-life applications.

In this study, we propose a novel multitasking learning framework named MTC-GA for glucose forecasting wherein each patient is treated as a separate task. The model framework with multiple personalized layers to extract the most significant features of each patient's BGL on its own. Moreover, it incorporates a shared layer that concatenates these extracted features from each patient facilitating the exchange of common patterns across tasks. Subsequently, task-specific layers distinguish between each patient's forecasted BGL with the whole process illustrated in Fig. 1. This architecture aims to combine personalization and generalization utilizing shared information while customizing forecasts for individual patients. While the novelty lies in integrating multiple advanced components into the framwork enhancing its forecasting capability. MTC-GA

is benchmarked against both personalized and generalized models with all of them being trained and validated using the OhioT1DM dataset. Thus, the key contributions for this work presented as follows:

- Proposing a novel framework incorporating multitasking architecture.
- Benchmarking against personalized and generalized approach to study the effect of the shared layer on the model.
- Benchmarking the proposed model against state-of-the-art methods with the highest accuracy reported in the literature for the Ohio dataset.
- Comprehensive multi-metric evaluation across a wide range of PHs.

The structure of this article is outlined as follows: Section II discusses the literature review with state-of-the-art models. Following this, Section III describes the methodology starting with the pre-processing steps followed by the model architecture itself, with the training process and hyper-parameters tuning. Subsequently, Section IV presents the results along with a discussion. Finally, Section VII presents the conclusions drawn from the paper.

II. RELATED WORK

To further substantiate the model's effectiveness, the proposed model was compared against state-of-the-art methods reported in the literature, particularly those with the highest accuracy when tested on the OhioT1DM dataset. Most of the studies employ personalized learning in their proposed solution, however, this is very computationally expensive in case of big data in real-case scenarios. Nevertheless, their developed architectures are very promising, forecasting with significant accuracy. A study proposed a basic LSTM personalized model for each patient resulting in an RMSE of 18.86 and 31.40 for PHs 30 and 60 minutes respectively [27]. While [28] focused on examining different imputing techniques with XGBoost model. Multiple features as carbohydrates intake, bolus insulin and heart rate were employed besides the historical blood glucose level. They finally trained 5 models each with different imputation technique achieving RMSE of almost 18.45 and MAE of 12.94 for 30 minutes prediction horizon. Both works involved only 6 patients from the 2018 version.

TABLE I
A SUMMARY OF STATE-OF-THE-ART MODELS, INCLUDING THE NUMBER OF PATIENTS USED IN THEIR STUDIES, THE FEATURES ADDED TO THE MODEL, THE SIGNIFICANT MODEL LAYER WITH THEIR TRAINING APPROACH, AND FINALLY THE PHs THE MODEL TRAINED ON.

Reference	Patients no.	Features	Model	Training approach	PHs	RMSE
[23]	6	BGL Carbohydrates Bolus insulin step counts	LSTM	Generalized	30,60	18.57, 30.32
[24]	12	BGL	LSTM	Generalized	30,60	18.26, 31.12
[25]	12	BGL Physical activity Carbohydrates Bolus insulin	Nested meta-learning LSTM/ MLP	Personalized	30,60	20.64, 32.98
[26]	12	BGL	Fuzzy c-mean SARIMA	Personalized	30, 45, 60, 75	20.13, 27.23, 31.96, 33.91
[27]	6	BGL	LSTM	Personalized	30,60	18.86, 31.4
[28]	6	BGL Bolus insulin Carbohydrates Heart rate	Ensemble XGBoost	Personalized	30	18.45
[29]	6	BGL	CNN-LSTM	Multitasking	30,60	19.79, 24.54
[30]	12	BGL	CNN-LSTM	Multitasking	30, 45, 60, 90, 120	18.8, 25.3, 31.8, 41.2, 47.2

Moreover, authors in [26] proposed a methodology of personalized forecasting model using SARIMA model for each patient individually. Their methodology involves grouping each patient's postprandial periods using fuzzy c-means to be fed to multiple models. Weighting the prediction of each model is the final forecasted value used for each patient. This ensemble model achieved an RMSE of 20.13, 27.23, 31.96 and 33.91 of prediction horizons of 30, 45, 60 and 75 minutes respectively. Other employed a more advanced methodology as presented in [25], authors suggested a nested meta-learning model with lag fusion for personalized BGL forecasting. The framework starts by finding the optimal history window with a compound lag fusion approach using nested ensemble learning. Three categories of learners were employed in the framework including non-stacking, stacking, and nested stacking. Each level's output acts as the next level's input forecasting at the end BGL with prediction of 30 and 60 mins. The model forecasted the OhioT1DM testing dataset with RMSE of 20.64 and 32.98 and MAE of 13.93 and 23.40 of prediction horizon of 30 and 60 mins respectively. Their limitation lies in filling null values in test sets using linear extrapolation which may induce artificial readings representing deceiving accuracy.

Conversely, a generalized approach takes place in multiple studies as in [23] who proposed a multi-layered LSTM model for blood glucose forecasting. They employed OhioT1DM dataset 2018 version which only uses 6 patients. Other features as carbohydrate intake from meals, the amount of bolus dose or infused insulin, exercise, and physical activeness are included in the study increasing in the accuracy of glucose level prediction. Pre-processing was done by Kalman smoothing, however, they trained the model with both raw and processed data. Thus, only the non-smoothed data results will be compared for a fair comparison. RMSE of both PHs of 30 and 60 minutes are 18.57 and 30.32 respectively. Furthermore, [24] was focusing on benchmarking multiple models as FNN, CNN, long short-

term memory network (LSTM), TCNN, and SAN. Each model was trained on multiple datasets including OhioT1DM, RT, DCLP5, and DCLP3 with different PHs. Focusing only of the RMSE of Ohio dataset, LSTM was their best model achieving RMSE of 18.26 and 31.12 while MAE of 12.15 and 21.83 of PH of 30 and 60 minutes, respectively.

As the best of our knowledge, multi-tasking approach is only presented in this study [29] which employs the multi-tasking learning with a traditional convolutional recurrent neural network called (MTCRNN) trained on 6 patients. The model was started by multiple shared layers to be fed to task-specific layers. Their limitation lies in the imputation method, by discarding samples with gaps of more than one hour resulting in losing much data. Despite this limitation, the model achieved an acceptable RMSE of 19.79 and 33.73 while MAE of 13.62 and 24.54 for PHs 30 and 60 minutes respectively. Subsequently, this work was continued in [30] to compare between multiple methodologies as single-task learning (STL) representing the personalized approach, sequential transfer learning (TL), multitasking learning and multi-tasking learning based on glycaemic variability (MTL-GV). MTL-GV is the same approach as the normal MTL with being clustered with a shared layer within each cluster based in the GV of patients. MTL was proved to achieve the lowest RMSE and MAE in all PHs except for 2 hours the MTL-GV achieved better. The MAE was 18.8, 25.3, 31.8, 41.2 and 47.2 for PHs 30, 45, 60, 90 and 120 minutes respectively. This study proves a significant low error across many PHs, however, their work lacks robust evaluation to measure the acceptability of the model clinically. These studies are summarized in table I, with their error being compared with the proposed methodology in the results section.

III. METHODOLOGY

A use-case of OhioT1DM dataset is employed, containing the blood glucose readings of 12 different patients with type

1 diabetes. It encompasses eight weeks collected every five minutes using Medtronic Enlite CGM sensors [31]. This section involves discussing the pre-processing steps done on this dataset, the model architecture developed for forecasting and the learning approach along with the tuned hyper-parameters.

A. Preliminary data handling

Starting with the pre-processing stage, dataset undergoes multiple steps to ensure uniformity across patients while curating the data for effective modeling as illustrated in Fig. 2.

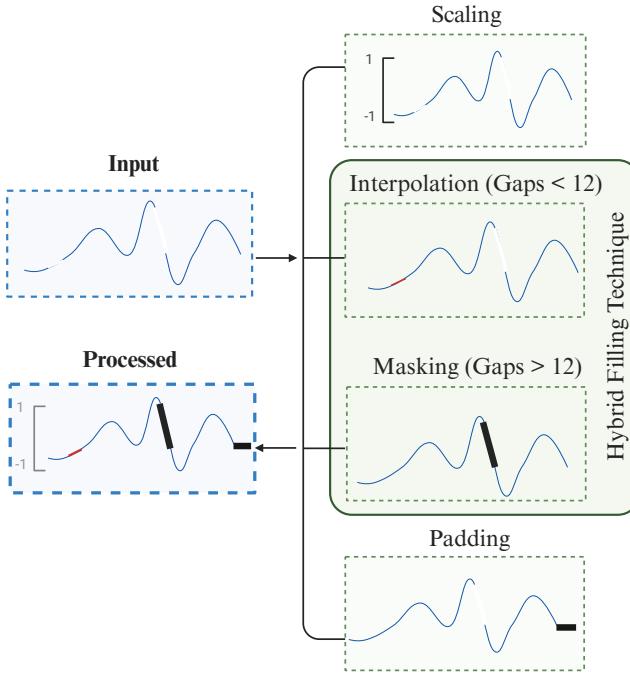


Fig. 2. Pre-processing pipeline starting with the input containing multiple issues and resulting in a processed series ready for the model. The processing steps involve scaling, interpolation for gaps smaller than 12 readings, masking null values for gaps longer than 12 readings and finally padding for uniform sequences.

The glucose readings are firstly resampled to a uniform time interval discovering the existence of null values with different gap lengths as presented in Fig. 3. Subsequently, the data is standardized to achieve zero mean and unit variance for better convergence of the model. Additionally, a right skewness is found in most of the patients; thus, log scaling is applied which also emphasizes any subtle variations in glucose trends. As presented in Fig. 4, the original patient's glucose series is skewed to value >1 , which is handled by log-scaling the values to be altered to a normal distribution.

Regarding the missing values discussed earlier, a hybrid strategy is employed to fill these values. A dual approach is configured and applied according to the gap length. If the gap length is smaller than 12 readings corresponding to one hour, this gap is filled using linear interpolation. Otherwise, if the gap length exceeded one hour, the values will be masked, explicitly handled during model training via masking mechanisms. Notably, imputation approaches that utilize both

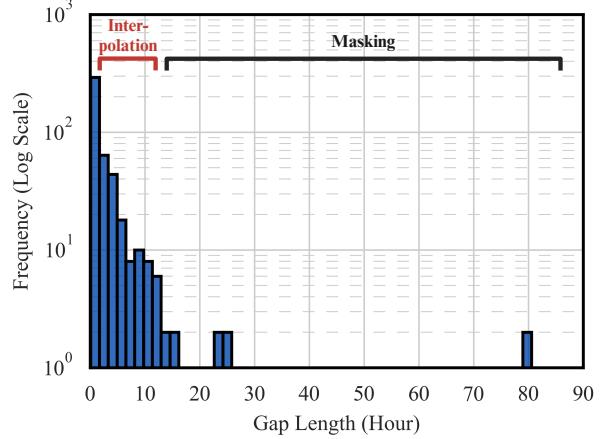


Fig. 3. A histogram of gap lengths for all 12 patients in the OhioT1DM dataset. The x-axis represents the gap length in hours while the y-axis (log-scaled) represents the number of gaps for each gap length. It also highlights the filling technique for each gap length, if length < 12 readings, interpolation is used, otherwise these values is masked with 0.

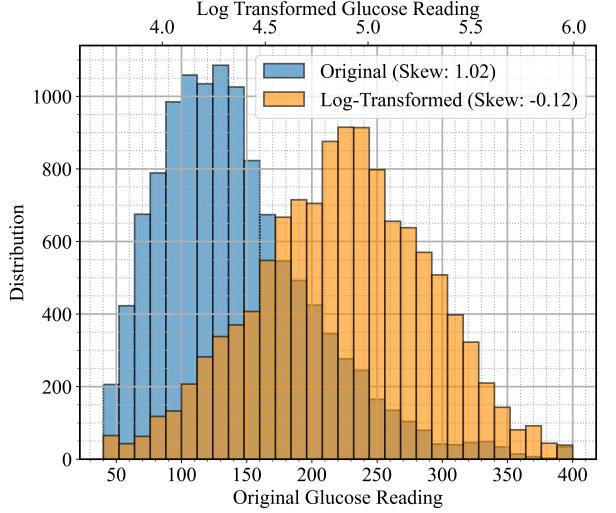


Fig. 4. The distribution of blood glucose values of patient 575 is illustrated in the blue histogram, which is transformed to be normally distributed using log scale presented in the orange histogram. The bottom x-axis is for the original readings, while the upper x-axis is for the log-transformed series.

border points of the data gap, the same as interpolation, are impractical in a real-world context as one boundary point lies in the future. Thus, the interpolation mechanism is altered to only use historical values moving in a forward direction only. Addressing the null gaps in the training dataset has been shown to enhance the accuracy of the model evaluated on the test dataset. Nonetheless, precautions are taken to guarantee that the test data stays unaltered, with gaps just masked rather than filled. This method eliminates the incorporation of synthetic data, therefore preventing biased results. Finally, due to the parallel training mechanism applied in the multitasking model, all input series must be the same length. Thus, sequences are padded to ensure uniform length and maintain compatibility with the architecture.

B. Multitasking model architecture

This study employs multitasking deep learning framework for forecasting blood glucose levels for multiple patients. The framework named MTC-GA denoted for all its components, which is multitasking(M), TCN(TC), GRU(G) and multihead attention(A). In general, multitasking models are designed to handle multiple related tasks simultaneously [32]. They aim to employ knowledge transfer mechanisms across tasks that improve generalization and efficiency. By doing so, these models can outperform single-task models in certain cases, such as having limited data for each task or when tasks have overlapping characteristics [33]. In contrast to this, single-task models can be better in case of sufficient data and complex specific characteristics [34]. Nevertheless, the greatest advantage of multitasking lies in reducing the computational and memory overhead [35]. In the time-series context, multitasking models are particularly used to learn both shared and task-specific temporal patterns for multiple related sequences, making them ideal for scenarios involving multiple interdependent time series. The most common form of multitasking time series forecasting model is starting with a shared layer followed by the task-specific layers with finally the output layer for all series [29], [30].

However, MTC-GA leverages personalized components at first to address individual-specific characteristics. In our case using OhioT1DM dataset, there are 12 patients, which corresponds to 12 different separate layers in the model. These layers allow the model to address each patient separately by learning representative features for each blood glucose sequence solely. Subsequently, all personalized layers are concatenated in one shared layer granting the learning of shared characteristics among all sequences. The model afterwards passes these features to distinct layers, acting as an output layer predicting each patient's series separately.

MTC-GA is illustrated in detail in Fig. 5 and its pseudocode is presented in Alg. 1. It involves adding multiple components for enhancing its capabilities in forecasting BGL for all patients. Firstly, the initial separate module is composed of a block of temporal convolutional followed by a GRU layer. The TCN block is repeated 2 times, each block involves a 1D convolutional layer for extracting temporal patterns followed by a Relu activation function and dropout rate. Another 1D CNN layer is added with the same configurations along with a residual connection between the output and first input allowing for smoother gradient flow. The first TCN block does not enable dilation in CNN to capture short dependencies while the second block uses a dilation rate of 2 to capture dependencies over longer periods. A simple down-sample CNN layer of kernel size 1 is added to ensure compatibility for the residual connection. Sequentially, a GRU layer is introduced which processes the TCN's output, learning patient-specific sequential patterns and further refining the temporal representations. A residual connection is similarly used here to combine the outputs of the TCN and GRU layers. Another concept is embedded to add an extra layer of complexity, is task-specific multi-head attention. This mechanism assigns adaptive importance weights to each time-step with its output being

Algorithm 1 MTC-GA Algorithm.

```

Require: Preprocessed data batches  $\{X_1, X_2, \dots, X_T\} \in \mathbb{R}^n$ , Number of tasks  $T$ 
Ensure: Final predictions  $Y \in \mathbb{R}^k$ 
1: Initialize shared_inputs  $\leftarrow []$ 
2: for each task  $i \in \{1, \dots, T\}$  do
3:    $X_{i_m}, M_i \leftarrow \text{Mask}(X_i)$ 
4:    $H_{t_i} \leftarrow \text{Conv1D}(X_{i_m}, k, s, p, d)$ 
5:    $H_{t_i} \leftarrow \text{ReLU}(H_{t_i})$ 
6:    $H_{t_i} \leftarrow \text{Dropout}(H_{t_i})$ 
7:    $H_{t_i} \leftarrow \text{Conv1D}(H_{t_i}, k, s, p, d)$ 
8:   if in_channels  $\neq$  out_channels then
9:      $X_{i_m} \leftarrow \text{Downsample}(X_{i_m})$ 
10:  end if
11:   $H_{t_i} \leftarrow \text{ReLU}(H_{t_i} + X_{i_m})$ 
12:   $H_{p_i} \leftarrow \text{GRU}_i(H_{t_i})$ 
13:   $H_{p_i} \leftarrow H_{p_i} + H_{t_i}$ 
14:   $Q, K, V \leftarrow H_{p_i}, H_{p_i}, H_{p_i}$ 
15:   $Q', K', V' \leftarrow \text{Linear}(Q), \text{Linear}(K), \text{Linear}(V)$ 
16:  Scores  $\leftarrow \text{Softmax}\left(\frac{Q'K'^T}{\sqrt{d_k}}\right)$ 
17:   $A_i \leftarrow \text{Scores} \cdot V'$ 
18:   $C_i \leftarrow \text{Mean}(A_i, \text{dim} = 1)$ 
19:  Append  $C_i$  to shared_inputs
20: end for
21:  $S \leftarrow \text{Stack}(\{C_1, \dots, C_T\}, \text{dim} = 1)$ 
22:  $H_s, \_ \leftarrow \text{SharedLSTM}(S)$ 
23:  $R_s \leftarrow \text{ResidualTransform}(S)$ 
24:  $H_s \leftarrow H_s + R_s$ 
25: Initialize personalized_outputs  $\leftarrow []$ 
26: for each task  $i \in \{1, \dots, T\}$  do
27:    $R_i \leftarrow \text{ResidualTransform}_i(H_s[:, i, :])$ 
28:    $P_i \leftarrow \text{ReLU}(\text{Dense}_i(H_s[:, i, :]))$ 
29:    $P_i \leftarrow \text{Dropout}(P_i)$ 
30:    $P_i \leftarrow P_i + R_i$ 
31:   Append  $P_i$  to personalized_outputs
32: end for
33:  $M \leftarrow \text{Concatenate}(\{P_1, \dots, P_T\}, \text{dim} = 1)$ 
34:  $Y \leftarrow \text{OutputLayer}(M)$ 
35: return  $Y$ 

```

averaged to produce a condensed representation highlighting the most informative features for each patient.

Moreover, the personalized outputs of all 12 submodels are concatenated and processed using a shared GRU layer with a residual connection between its output and the stacked output. This approach captures common features among all BGL data facilitating inter-patient knowledge transfer. For generating the output, the model is finalized with a personalized dense layer at the end with a residual connection to be concatenated to an output layer generating the prediction.

C. Training process

To evaluate the effectiveness of multitasking in parallel time-series forecasting, personalized and generalized modeling approaches with the same architecture are implemented. Both models utilize a single branch of the multi-tasking architecture

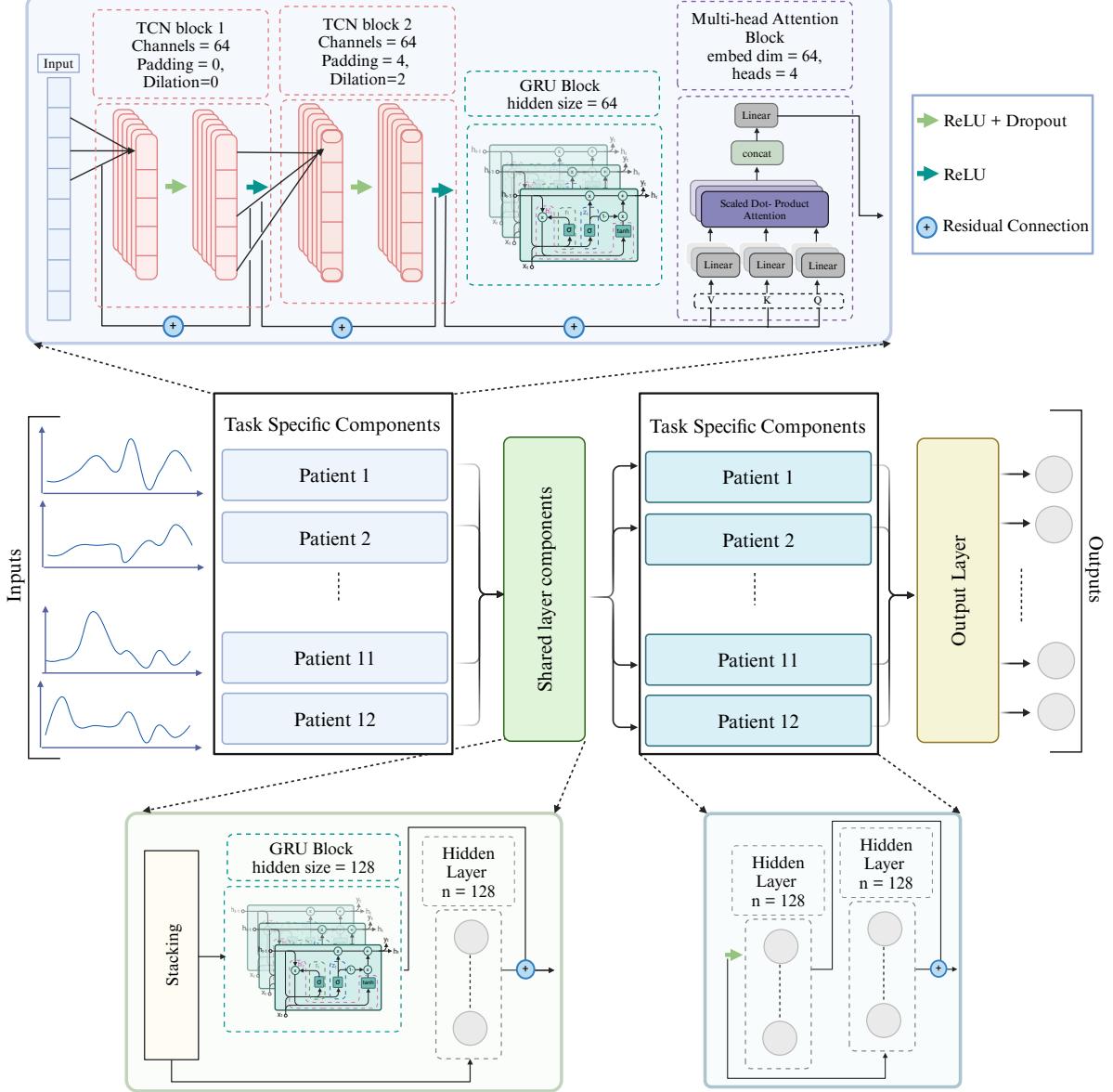


Fig. 5. MTC-GA architecture starts with taking multiple inputs to task-specific layers incorporating two TCN blocks and a GRU block, followed by a multi-head attention block to be concatenated and passed to shared layer components and returned again to task-specific block of hidden layers and finally passed to output layer, forecasting each patient's BGL solely.

for either single time-series input (personalized model) or stacked time-series inputs from multiple patients (generalized model). The personalized model focuses on modeling the glucose dynamics of an individual patient, thus, it excels in capturing fine-grained individual patterns. This makes it more complex and requires more data for effective training. In contrast to this, the generalized model focuses on the inter-patient relationships which only learn the shared features among patients. This maintains the computational complexity and may increase efficiency in case of limited individual readings or noisy datasets. The complexity of the three models is presented in Table II providing the number of trainable parameters in each one. The personalized approach is the highest complexity among all, with the number of

trainable parameters for each patient equal 170,177, thus the final complexity after being multiplied by 12 model equals 2,042,124. While the multitasking and the generalized model has 1,272,460 and 170,892 number of trainable parameters respectively. The evaluation will be focused on benchmarking the multitasking, personalized, and generalized approach to determine the optimal model strategy for real-world application.

D. Hyper-parameters tuning

To ensure a fair comparison between the three learning approaches, the hyperparameters of all models are fine-tuned to be identical as presented in Table III-D. The training employs the L1 loss function optimized by Adam with initial

TABLE II

NUMBER OF TRAINABLE PARAMETERS OF THE THREE APPROACHES REPRESENTING THE COMPLEXITY OF EACH MODEL.

Approach	Personalized (x12)	Multitasking	Generalized
No. of parameters	2,042,124	1,272,460	170,892
Training time (min)	264	68	23
Inference time (s)	40.68	4.81	2.45

TABLE III

HYPER-PARAMETER CONFIGURATION FOR THE THREE LEARNING APPROACHES

Hyper-parameter	Value
Loss Function	L1 Loss
Optimizer	Adam
Initial Learning Rate	$1e - 5$
Number of Epochs	150
Batch Size	64
Mask Value for Missing Data	0
Window Size (mins)	60
PHs (mins)	15, 30, 45, 60, 75, 90, 120
Number of Runs for each model	5
Number of TCN Blocks	2
Multi-Head Embedded Dimension	64
No. of Multi-Head Attentions	4
GRU Hidden Sizes	64, 128

learning rate of $1e - 5$. Each model was run five times each for 150 epochs with batch size of 64, and the mean and standard error of the results across these runs were calculated to account for variability in training and to ensure robust performance evaluation. Regarding the time-series processing parameters, masking the missing values is handled to be filled with 0, making sure that 0 is not included in the data range after being log-scaled of the standardized values. After trying multiple window sizes for input sequences, it is determined to be fixed at 12 time steps corresponding to one hour. Moreover, the model is being evaluated across multiple PHs of 3, 6, 9, 12, 15 and 18 time steps equivalent to 15, 30, 45, 60, 90 and 120 minutes to enable a comprehensive assessment of their forecasting performance over temporal scales. This unified hyperparameter configuration ensures that performance differences between the models are attributable to their architectural differences rather than variations in training or optimization parameters.

IV. RESULTS & DISCUSSION

Loss curves of all three approaches are presented in Fig. 6, showing the convergence of both personalized and MTC-GA contrary to the generalized model. Moreover, all approaches are evaluated using mean absolute error (MAE) and root mean square error (RMSE). Table S1 presents the error of all models with each patient across multiple PHs. Each approach exhibits a different behavior among each PH. Visualizing the mean of each technique across each PH highlights the performance gap between the generalized model and both the personalized and multi-tasking models in the error and its stability among each run as illustrated in Fig. 7.

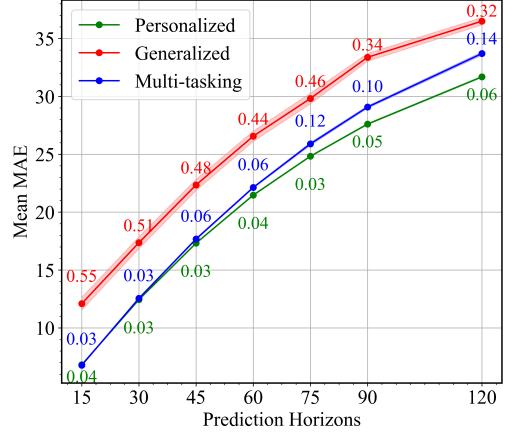


Fig. 7. The MAE mean of all 5 runs of personalized, generalized and multi-tasking approach with , with the standard deviations (shaded areas) annotated for clarity across all PHs.

By interpreting the mean error of each approach in Fig. 7, both personalized and multi-tasking models have nearly the same MAE across PHs of 15 and 30 mins, however, the difference of error between both kept increasing as the PH increases. This highlights the personalized model's ability to capture fine-grained patient-specific patterns especially in long-term forecasts. Regarding the generalized, it exhibits a significant gap compared to other approaches with a high variability across different runs. This proves the inability of the generalized model to rely only on shared patterns across patients. Nevertheless, looking closer at each patient's forecasting error, MTC-GA outperforms the personalized approach in multiple use cases as highlighted in the heatmap in Fig. IV and summarized in Table S1.

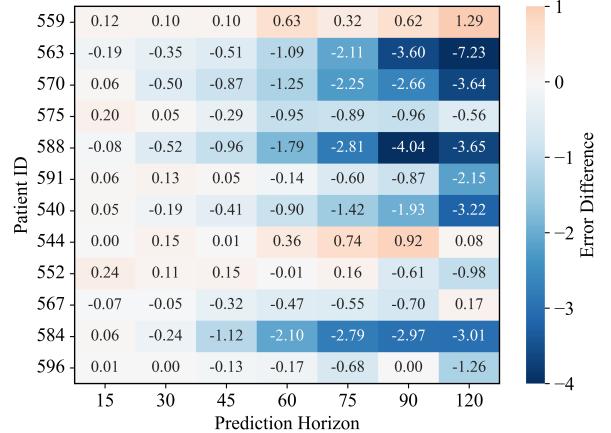


Fig. 8. A heatmap showing the difference between the personalized and multi-tasking MAE for each patient across all PHs. Positive value indicates a higher personalized MAE while negative values is the opposite indicating a lower personalized MAE.

The figure shows the difference in error between both approaches in all patients. As the color gets darker in the blue shade, it indicates a higher MAE in the multi-tasking approach compared to the personalized approach. There are three patients out of 12 that got better results in the multitasking approach, while three patients can be considered nearly the same with a small difference. The other six patients perform

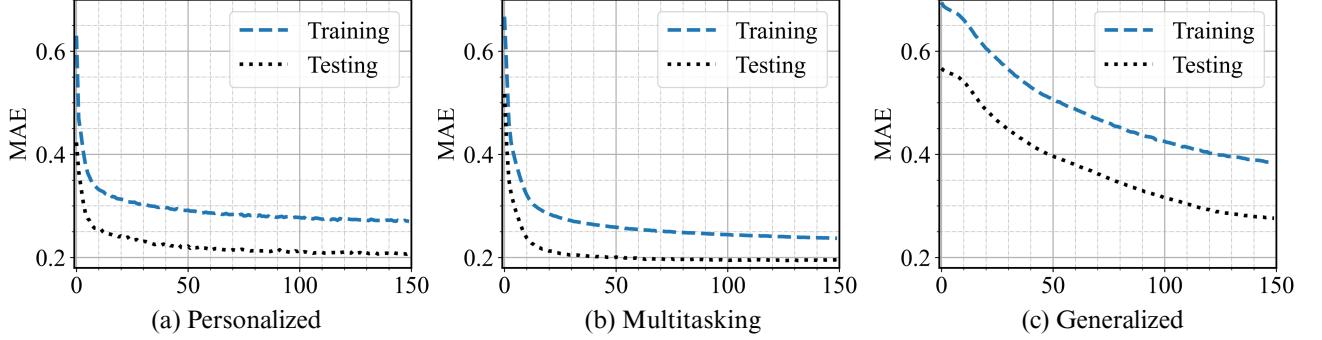


Fig. 6. The MAE loss curves of the three approaches: (a) the personalized model, (b) the multitasking model (MTC-GA), and (c) the generalized model.

worse in the multitasking approach. This distinct behavior among patients can be explained by identifying the distance between their glucose patterns. The dendrogram presented in Fig. 9 shows a very small distance among patients 559, 544 and 567 who follow the same behavior in their error of predicted values by MTC-GA.

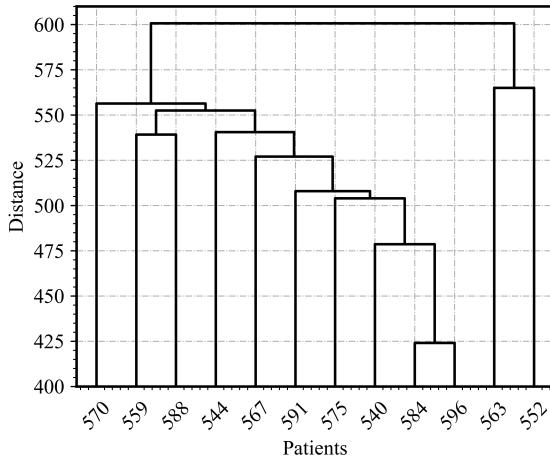


Fig. 9. Dendrogram visualizing the distance between all patients' BGL series. The x-axis represents the patient code while the y-axis is the distance value.

Furthermore, a clinical evaluation method called clarke error grid (CEG) was obtained as presented in Fig. 10. Specifically, it assesses the prediction errors by categorizing errors into different zones based on their clinical significance. The presented CEG shows 0% of data falls in zone E at PHs of 15, 30 and 45 mins. While having a very small fraction of data in other PHs with the highest error of 9e – 2% in the 2 hour PH. Regarding zone D which indicates the missed hypoglycemic or hyperglycemic events, it starts with 0.26% in 15 mins PH and kept increasing until it was 5.6% in the highest PH. Also, Zone C contains a very small number of predictions, with the highest score being 0.25% considered a very low number. Finally, Zone A and Zone B contain the highest percentage among all PHs presenting an overall accurate prediction of the model. Thus, the proposed model can be considered a clinically accepted model across PHs of 15, 30 and 45 mins. Although the model has a very small fraction of error in zone E, unfortunately it is considered a dangerous zone, and

the model may be considered less clinically acceptable in higher PHs. Moreover, Zone A and Zone B are considered the clinically accurate or benign errors, in contrast, Zones C, D and E demonstrated the more significant errors. Thus, each category was summed and presented in Table IV.

For more evaluation metrics, Table IV shows all other metrics obtained on the MTC-GA's prediction. The R^2 score

TABLE IV
EVALUATION METRICS OBTAINED FOR THE PROPOSED MODEL AT DIFFERENT PHs RANGING FROM 15 TO 120 MINUTES. IT PRESENTS THE R^2 SCORE, MADP, MAE, RMSE AND THE CEG ZONES BINNED INTO 2 CATEGORIES, ZONE A+B AND ZONE C+D+E.

Eval Metric	PHs (mins)						
	15	30	45	60	75	90	120
R^2 Score	0.95	0.85	0.73	0.61	0.48	0.36	0.17
MADP %	4.89	9.01	12.75	15.98	18.82	21.30	24.97
A+B%	99.72	98.83	97.57	96.80	96.17	95.40	94.04
C+D+E%	0.27	1.16	2.42	3.19	3.82	4.59	5.95
MAE	6.77	12.55	17.68	22.13	25.90	29.08	33.70
RMSE	10.36	17.95	24.49	29.93	34.52	38.12	43.46

which is also named the coefficient of determination, indicates a great score at the first PHs while showing a decreasing trend as the PH increases. The 15, 30 and 45 mins PH can be considered an acceptable R^2 score from 0.95 to 0.73. Although the R^2 score indicates how much of the variability in the actual data is captured by the model, it is still misleading for a nonlinear model, as in our case. It normally assumes a linear relationship and it may not accurately reflect goodness-of-fit [36]. Furthermore, the mean absolute difference percentage (MADP) starts with only 4.8% of error and keeps increasing until it reaches 24.97% of error which still demonstrates a strong performance even in high PH. It is found that the MADP increases with factor of almost 1.36 which indicates that the model's performance degrades non-linearly but predictably as the PH increases, reflecting good model stability. Moreover, this performance indicates a higher ability to predict at longer PHs due to its reduced degradation rate. Specifically, the error grows as the PH increases normally, however, the error growth between smaller and larger PHs is relatively gradual with smaller intervals in MADP values between longer horizons as 4.12, 3.74, 3.23, 2.84 and 2.4.

Compared to state-of-the-art models, Table V represents the comparison of MAE and RMSE for BGL forecasting across all presented PHs. MTC-GA model consistently demonstrates

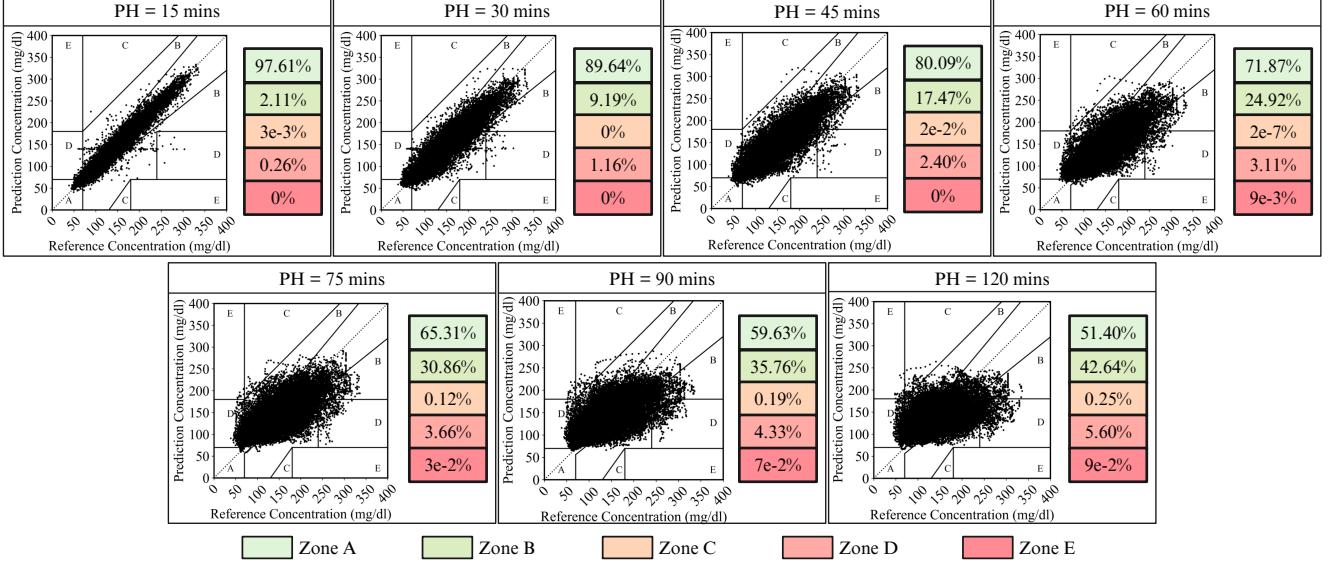


Fig. 10. The CEG evaluating the clinical accuracy of the glucose prediction models of all PHs. The grid is divided into five zones (A, B, C, D, and E) that represent the degree of agreement between predicted glucose values and reference glucose values. Zone A is the most clinically safe zone, while it ends with Zone E being the most dangerous zone.

TABLE V

THE MAE AND RMSE OF MTC-GA COMPARED TO THE STATE-OF-THE-ART MODELS IN LITERATURE, WITH MTC-GA ACHIEVING THE LOWEST ERROR AMONG ALMOST ALL MODELS.

Ref.		[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]	MTC-GA
PHs (mins)	Eval metric	-	-	-	-	-	-	-	-	6.77 ± 0.03
15	MAE	-	-	-	-	-	-	-	-	6.77 ± 0.03
	RMSE	-	-	-	-	-	-	-	-	10.36 ± 0.04
30	MAE	-	12.15 ± 0.14	13.93 ± 0.51	-	-	-	13.62 ± 0.05	13.2 ± 1.60	12.55 ± 0.03
	RMSE	18.57	18.26± 0.25	20.64 ± 0.63	20.13	18.86 ± 1.79	18.45 ± 2.55	19.79 ± 0.06	18.8 ± 2.30	17.95 ± 0.05
45	MAE	-	-	-	-	-	-	-	18.2 ± 2.20	17.68 ± 0.06
	RMSE	-	-	-	27.23	-	-	-	25.3 ± 2.90	24.49 ± 0.07
60	MAE	-	21.83 ± 0.23	23.4 ± 0.77	-	-	-	24.54 ± 0.08	23.4 ± 3.00	22.13 ± 0.06
	RMSE	30.32	31.12 ± 0.35	32.98 ± 0.73	31.96	31.4 ± 2.07	-	33.73 ± 0.24	31.8 ± 3.90	29.93 ± 0.08
75	MAE	-	-	-	-	-	-	-	-	25.90 ± 0.12
	RMSE	-	-	-	33.91	-	-	-	-	34.52 ± 0.16
90	MAE	-	-	-	-	-	-	-	31.1 ± 3.70	29.08 ± 0.10
	RMSE	-	-	-	-	-	-	-	41.2 ± 4.50	38.12 ± 0.14
120	MAE	-	-	-	-	-	-	-	36.5 ± 3.80	33.70 ± 0.14
	RMSE	-	-	-	-	-	-	-	47.2 ± 4.60	43.46 ± 0.12

lower RMSE and MAE values compared to almost all other models. There is no discussed model in the literature, trained on PH of 15 mins. Thus, starting with PH of 30 mins, MTC-GA achieves an RMSE of 17.95 outperforming all others which are ranging between 18.26 and 20.64, however, [24] performs better with difference of MAE equals 0.40. This suggests that MTC-GA is better at minimizing the impact of large deviations and handling large errors, which is crucial in our application; on the other hand, it slightly sacrifices a bit of overall error magnitude. Regarding the PH of 45 mins, only [30] calculates its error with a higher MAE and RMSE

than MTC-GA with a difference of almost one. At 60 mins PH, [24] has the same performance of better MAE and worse RMSE than the proposed MTC-GA while it still outperforms all others. For 75 mins PH, [26] performs better than MTC-GA, however, MTC-GA demonstrates superior forecasting accuracy in all other PHs compared to this model particularly, with a great difference between them of -2.18, -2.74, -2.03, and +0.61 in PHs 30, 45, 60 and 75 mins respectively. Regarding the model stability across multiple runs, MTC-GA is the most stable model, ranging between 0.03 and 0.14 across all PHs while the lowest std of all other models was starting

from 0.14 and reaching till 4.60 indicating a very unstable model.

V. CONCLUSION

The proposed MTC-GA demonstrates a robust balance between personalized and generalized approaches offering a practical solution for multi-patient glucose level forecasting. While the generalized model is scalable and efficient in terms of complexity, its accuracy diminishes particularly for long-term predictions making it not applicable. Conversely, the personalized model achieves a comparably low MAE across all PHs excelling in capturing individual glucose dynamics but requiring substantial patient-specific data which limits its applicability in real scenarios. MTC-GA effectively leverages shared patterns across patients while preserving individual-level accuracy, maintaining competitive MAE values and reduced variability particularly in shorter PHs. Although it slightly underperforms compared to personalized models in high PHs, it is significantly lower in complexity by 40% showcasing its ability to deliver reliable and efficient forecasting for real-world multi-patient scenarios. Moreover, the proposed MTC-GA outperformed all the state-of-the-art models even in the high PHs. This highlights its suitability as a scalable and accurate compromise, capable of addressing the limitations of existing models in both precision and applicability.

VI. DATA AVAILABILITY

The OhioT1DM dataset used in this study is not publicly available. Access to this dataset can be obtained by researchers through a Data Use Agreement (DUA) with Ohio University. Researchers can request a DUA via <http://smarthealth.cs.ohio.edu/OhioT1DM-dataset.html>. Detailed information about the dataset is provided in the publication by Marling, C. et al., OhioT1DM Dataset for Blood Glucose Level Prediction, available at <https://pubmed.ncbi.nlm.nih.gov/33584164/>, DOI: 10.2196/16662. The code for the implemented algorithms used in this study is publicly available on GitHub and can be accessed via the following link: <https://github.com/Mennatullah/Glucose-forecasting-using-multitasking>.

REFERENCES

- [1] T. N. Gia, M. Ali, I. B. Dhaou, A. M. Rahmani, T. Westerlund, P. Liljeberg, and H. Tenhunen, "Iot-based continuous glucose monitoring system: A feasibility study," *Procedia Computer Science*, vol. 109, pp. 327–334, 2017, 8th International Conference on Ambient Systems, Networks and Technologies, ANT-2017 and the 7th International Conference on Sustainable Energy Information Technology, SEIT 2017, 16-19 May 2017, Madeira, Portugal. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050917310281>
- [2] Y. Liu, M. Kang, S. Gao, C. Zhang, Y. Liu, S. Li, Y. Qi, A. Nathan, W. Xu, C. Tang, E. Occhipinti, M. Yusufu, N. Wang, W. Bai, and L. Occhipinti, "Diagnosis of multiple fundus disorders amidst a scarcity of medical experts via self-supervised machine learning," *IEEE Internet of Things Journal*, vol. 12, no. 1, pp. 224–235, 2025.
- [3] A. S. Duggal, P. K. Malik, H. Bedi, R. Roges, and R. Singh, "Emerging technologies in the field of smart monitoring healthcare system with cardiac disease," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–6.
- [4] T. L. Hong, Y. Dave, A. Khant, L. Verma, M. Chauhan, and S. Parthasarathy, "The future of personalized medicine and internet of things reshaping healthcare treatment plans and patient experiences," *Journal of Intelligent Systems and Internet of Things*, vol. 14, no. 1, p. 230 – 240, 2025. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85204046297&doi=10.54216%2fJISIoT.140118&partnerID=40&md5=2d2cf36e7838dfa9ea6e67bc344c59f>
- [5] H. Manoharan, D. Jayaseelan, and S. Appu, "A comparative study on continuous glucose monitoring devices for managing diabetes mellitus," *Revue d'Intelligence Artificielle*, vol. 37, no. 5, p. 1351 – 1360, 2023. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85178288930&doi=10.18280%2fria.370528&partnerID=40&md5=dd038976a60f4ec6d0443c2440328632>
- [6] R. W. Beck, T. Riddlesworth, K. Ruedy, A. Ahmann, R. Bergenstal, S. Haller, C. Kollman, D. Kruger, J. B. McGill, W. Polonsky, E. Toschi, H. Wolpert, D. Price, and D. S. Group, "Effect of continuous glucose monitoring on glycemic control in adults with type 1 diabetes using insulin injections: The diamond randomized clinical trial," *JAMA*, vol. 317, no. 4, pp. 371–378, 2017. [Online]. Available: <https://doi.org/10.1001/jama.2016.19975>
- [7] L. P. Ying, O. X. Yin, O. W. Quan et al., "Continuous glucose monitoring data for artificial intelligence-based predictive glycemic event: A potential aspect for diabetic care," *International Journal of Diabetes in Developing Countries*, 2024. [Online]. Available: <https://doi.org/10.1007/s13410-024-01349-x>
- [8] L. Hölzen, B. Schultes, S. M. Meyhöfer, and S. Meyhöfer, "Hypoglycemia unawareness—a review on pathophysiology and clinical implications," *Biomedicines*, vol. 12, no. 2, 2024. [Online]. Available: <https://www.mdpi.com/2227-9059/12/2/391>
- [9] E. Karslioglu French, A. C. Donihi, and M. T. Korytkowski, "Diabetic ketoacidosis and hyperosmolar hyperglycemic syndrome: review of acute decompensated diabetes in adult patients," *BMJ*, vol. 365, 2019. [Online]. Available: <https://www.bmjjournals.org/content/365/bmj.l1114>
- [10] T. Yang, F. Qi, F. Guo et al., "An update on chronic complications of diabetes mellitus: from molecular mechanisms to therapeutic strategies with a focus on metabolic memory," *Molecular Medicine*, vol. 30, p. 71, 2024. [Online]. Available: <https://doi.org/10.1186/s10020-024-00824-9>
- [11] H. Nemat, H. Khadem, J. Elliott et al., "Data-driven blood glucose level prediction in type 1 diabetes: a comprehensive comparative analysis," *Scientific Reports*, vol. 14, p. 21863, 2024. [Online]. Available: <https://doi.org/10.1038/s41598-024-70277-x>
- [12] H. Deng, L. Zhang, Y. Xie, and S. Mo, "Research on estimation of blood glucose based on ppg and deep neural networks," *IOP Conference Series: Earth and Environmental Science*, vol. 693, no. 1, p. 012046, mar 2021. [Online]. Available: <https://doi.org/10.1088/1755-1315/693/1/012046>
- [13] G. Alfiari, M. Syafrudin, M. Anshari, F. Benes, F. T. D. Atmaji, I. Fahrurrozi et al., "Blood glucose prediction model for type 1 diabetes based on artificial neural network with time-domain features," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1586–1599, 2020. [Online]. Available: <https://doi.org/10.1016/j.bbce.2020.10.004>
- [14] J. B. Ali, T. Hamdi, N. Fnaiech, V. Di Costanzo, F. Fnaiech, and J. M. Ginoux, "Continuous blood glucose level prediction of type 1 diabetes based on artificial neural network," *Biocybernetics and Biomedical Engineering*, 2020.
- [15] Y. Deng, L. Lu, L. Aponte, A. M. Angelidi, V. Novak, G. E. Karmiadakis et al., "Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients," *NPJ Digital Medicine*, vol. 4, no. 1, p. 109, 2021. [Online]. Available: <https://doi.org/10.1038/s41746-021-00480-x>
- [16] W. Seo, S. W. Park, N. Kim, S. M. Jin, and S. M. Park, "A personalized blood glucose level prediction model with a fine-tuning strategy: A proof-of-concept study," *Computer Methods and Programs in Biomedicine*, vol. 211, p. 106424, 2021. [Online]. Available: <https://doi.org/10.1016/j.cmpb.2021.106424>
- [17] J. Xie and Q. Wang, "Benchmarking machine learning algorithms on blood glucose prediction for type i diabetes in comparison with classical time-series models," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 11, pp. 3101–3124, 2020. [Online]. Available: <https://doi.org/10.1109/TBME.2020.2975959>
- [18] S. Bhargav, S. Kaushik, V. Dutt et al., "Temporal convolutional networks involving multi-patient approach for blood glucose level predictions," in *2021 International Conference on Computational Performance Evaluation (ComPE)*. IEEE, 2021, pp. 288–294.
- [19] H. Nemat, H. Khadem, M. R. Eissa, J. Elliott, and M. Benaiissa, "Blood glucose level prediction: advanced deep ensemble learning

- approach,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 6, pp. 2758–2769, 2022. [Online]. Available: <https://doi.org/10.1109/JBHI.2022.3144870>
- [20] R. Cui, C. Hettiarachchi, C. J. Nolan, E. Daskalaki, and H. Suominen, “Personalised short-term glucose prediction via recurrent self-attention network,” in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2021, pp. 154–159.
- [21] T. Zhu, K. Li, P. Herrero, and P. Georgiou, “Personalized blood glucose prediction for type 1 diabetes using evidential deep learning and meta-learning,” *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 1, pp. 193–204, 2022. [Online]. Available: <https://doi.org/10.1109/TBME.2022.3187703>
- [22] T. Zhu, L. Kuang, J. Daniels, P. Herrero, K. Li, and P. Georgiou, “Iotm-enabled real-time blood glucose prediction with deep learning and edge computing,” *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 3706–3719, 2022. [Online]. Available: <https://doi.org/10.1109/JIOT.2022.3143375>
- [23] M. F. Rabby, Y. Tu, M. I. Hossen *et al.*, “Stacked lstm based deep recurrent neural network with kalman smoothing for blood glucose prediction,” *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, pp. 1–12, 2021.
- [24] S. Ghimire, T. Celik, M. Gerdes, and C. W. Omlin, “Deep learning for blood glucose level prediction: How well do models generalize across different data sets?” *PLoS One*, vol. 19, no. 9, p. e0310801, Sep 2024.
- [25] H. Khadem, H. Nemat, J. Elliott, and M. Benaissa, “Blood glucose level time series forecasting: Nested deep ensemble learning lag fusion,” *Bioengineering*, vol. 10, no. 4, 2023. [Online]. Available: <https://www.mdpi.com/2306-5354/10/4/487>
- [26] F. Prendin, J. L. Díez, S. Del Favero, G. Sparacino, A. Facchinetto, and J. Bondia, “Assessment of seasonal stochastic local models for glucose prediction without meal size information under free-living conditions,” *Sensors (Basel, Switzerland)*, vol. 22, no. 22, p. 8682, 2022.
- [27] J. Martinsson, A. Schliep, B. Eliasson *et al.*, “Blood glucose prediction with variance estimation using recurrent neural networks,” *Journal of Healthcare Informatics Research*, vol. 4, pp. 1–18, 2020.
- [28] J. Jeon, P. J. Leimbigler, G. Baruah, M. H. Li, Y. Fossat, and A. J. Whitehead, “Predicting glycaemia in type 1 diabetes patients: Experiments in feature engineering and data imputation,” *Journal of Healthcare Informatics Research*, vol. 4, no. 1, pp. 71–90, 2019.
- [29] J. Daniels, P. Herrero, and P. Georgiou, “Personalised glucose prediction via deep multitask networks,” in *KDH@ECAI*, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:221793894>
- [30] ———, “A multitask learning approach to personalized blood glucose prediction,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 436–445, 2022.
- [31] C. Marling and R. Bunescu, “The ohio t1dm dataset for blood glucose level prediction: Update 2020,” in *CEUR Workshop Proceedings*, vol. 2675. NIH Public Access, 2020, p. 71.
- [32] G. Ye, Z. Lu, D. Zhang, F. Zhou, R. Song, L. Shu, Y. Pu, and Y. Chen, “Learning unified model for sleep health monitoring,” in *2024 8th International Conference on Communication and Information Systems (ICCIS)*, 2024, pp. 27–32.
- [33] Y. Zhang and Q. Yang, “A survey on multi-task learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 12, pp. 5586–5609, 2022.
- [34] G. Rotman and R. Reichart, “Multi-task active learning for pre-trained transformer-based models,” *Transactions of the Association for Computational Linguistics*, vol. 10, pp. 1209–1228, 11 2022. [Online]. Available: https://doi.org/10.1162/tacl_a_00515
- [35] S. Miraliev, S. Abdigapporov, V. Kakani, and H. Kim, “Real-time memory efficient multitask learning model for autonomous driving,” *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 247–258, 2024.
- [36] V. Plevris, G. Solorzano, N. P. Bakas, and M. E. A. Ben Seghier, “Investigation of performance metrics in regression analysis and machine learning-based prediction models,” in *8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2022)*. European Community on Computational Methods in Applied Sciences, 2022.