Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 4 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i04.60113

PAPER

Generalizing CGM Sensor-Based Glucose Prediction across Age Cohorts Using LSTM Models: An In Silico Study with the UVA/Padova T1DMS Simulator

Saleh I. Alzahrani (✉)

College of Engineering, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

sialzahrani@iau.edu.sa

ABSTRACT

Accurate glucose prediction is necessary in enhancing insulin therapy and preventing harmful blood-sugar spikes in individuals with type 1 diabetes (T1D). Although deep learning models seem promising for glucose forecasting, it is unclear how they perform across various age groups that exhibit different metabolism profiles. This paper compares the performance of the long short-term memory (LSTM) models across age groups using simulated data from the UVA/Padova T1D Metabolic Simulator (T1DMS). Cohort-specific models achieved high within-cohort performance (MAE < 2 mg/dL, r > 0.99, R^2 > 0.99), indicating precise modeling of glucose–insulin dynamics within each group. Nevertheless, predictive accuracy decreased when models were applied across cohorts, especially when LSTM networks trained on adults were tested on younger groups, demonstrating physiological variability between ages in insulin sensitivity and glucose kinetics. Models trained on younger groups performed better on older populations, suggesting that a broader range of metabolic variation underlies increased adaptability. This study is the first to use the FDA-approved UVA/Padova T1DMS simulator to systematically assess age-dependent generalization in LSTM-based glucose prediction, offering a unique reproducible framework for developing adaptive e-health and closed-loop insulin systems. Incorporating age-relevant physiological heterogeneity and adaptive modeling paradigms could help develop stronger, patient-specific glucose forecasting systems for safer and more effective diabetes management.

KEYWORDS

continuous glucose monitoring, long short-term memory (LSTM), type 1 diabetes (T1D), glucose prediction, closed-loop insulin systems

1 INTRODUCTION

Diabetes mellitus (DM) is a metabolic disorder characterized by elevated blood glucose levels that, if inadequately managed, can lead to severe health

Alzahrani, S. I. (2026). Generalizing CGM Sensor-Based Glucose Prediction across Age Cohorts Using LSTM Models: An In Silico Study with the UVA/Padova T1DMS Simulator. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(4), pp. 76–93. https://doi.org/10.3991/ijoe.v22i04.60113

Article submitted 2025-12-16. Revision uploaded 2026-01-02. Final acceptance 2026-01-05.

complications including microangiopathy, retinopathy, nephropathy, and cardiovascular disease [1]–[3]. Type 1 diabetes (T1D), formerly called juvenile diabetes or insulin-dependent diabetes, results from the autoimmune destruction of β-cells within the islets of Langerhans. Although less than 10% of diabetic cases are classified as T1D, this proportion is increasing worldwide [4]. The resulting insulin deficiency requires lifelong exogenous insulin therapy and careful glycemic management to maintain blood glucose levels within a target range and to reduce the risk of acute and chronic complications associated with hypoglycemia and hyperglycemia [5], [6].

Accurate prognostication and diagnosis of diabetes are important tools for optimizing therapeutic interventions and preventing both acute and chronic complications [7], [8]. Recent technological progress, such as continuous glucose monitoring (CGM) and the insulin pump, has completely changed the way metabolic diseases are handled via the provision of real-time glycemic data and automated insulin dosing [9]. However, accurate short-term glucose prediction remains challenging, as the relationships between glucose and insulin are highly nonlinear and idiosyncratic [10], [11]. In this context, forecasting models can optimize the process of glycemia control and potentially prevent disease progression to advanced stages.

In recent years, various machine learning-based predictive models have been developed to enable the initial diagnosis and risk identification of diabetes [12]–[16]. Various approaches, including deep learning models such as long short-term memory (LSTM) networks, have demonstrated significant potential in modeling long-term dependencies and nonlinear interactions across sequential time steps. Consequently, these architectures have proven themselves to be satisfactory in predicting diabetes [17]–[20]. However, many studies rely solely on internal validation, reducing their generalizability and therapeutic value [21]. Moreover, demographic and physiological differences based on age—such as lifestyle behavior, comorbidities, body composition, and hormonal regulation—often reduce the predictive accuracy of models trained on heterogeneous age groups.

In this study, we investigated how LSTM-based glucose prediction models generalize across different age cohorts. While numerous studies have applied LSTM models for diabetes prediction, most rely on limited clinical datasets and evaluate their performance only within single cohorts. In contrast, this work makes two key contributions. First, we employed the FDA-approved T1DMS to generate controlled and reproducible in silico datasets for children, adolescents, and adults under consistent conditions, such as meal timing and insulin delivery, that are difficult to control in clinical settings. This simulation-based framework enables researchers to systematically examine cross-cohort models while avoiding the ethical and logistical challenges of clinical trials. Second, we investigated cross-cohort generalization by training LSTM models on one age group and validating them on others to evaluate their effectiveness and transferability, thereby guiding the development of more adaptive glucose prediction frameworks.

The primary objective of this work is to determine whether a single LSTM model trained on a single age cohort can accurately predict glucose trajectories in other age cohorts, or whether a model that has been trained and customized to a specific age cohort is necessary to maintain predictive accuracy. It is hypothesized that the findings of this study will contribute to developing robust, age-independent glucose prediction systems that have the potential to support personalized diabetes management and advance next-generation artificial pancreas systems. From a modeling perspective, age-related physiological differences can be interpreted as a form of structured uncertainty in glucose–insulin dynamics. This view motivates the investigation of simple adaptive mechanisms that can preserve prediction performance when models are applied across different age cohorts.

The subsequent sections of this paper are organized as follows: Section 2 describes the data generation process using the UVA/Padova T1DMS, the preprocessing steps, and the modeling framework, including the architecture, training, and evaluation procedures of the LSTM networks. Section 3 presents the results of within- and cross-cohort analyses, along with quantitative and graphical evaluations of model performance. Section 4 discusses the physiological interpretation of the findings, the implications of age-dependent model generalization, and potential directions for improving glucose prediction accuracy. Finally, Section 5 provides the main conclusions and outlines future research perspectives.

2 MATERIALS AND METHODS

2.1 Data description

In this study, the University of Virginia/Padova (UVA/Padova) T1DMS (version 3.2, The Epsilon Group, Charlottesville, Virginia) was used to generate in silico subjects with T1D. The UVA/Padova T1DMS is the first in silico diabetes model approved by the U.S. Food and Drug Administration (FDA) in 2008 as a preclinical testing tool of closed-loop algorithms [22], [23]. The simulated subjects were categorized into three age groups: children (N = 11, 5.2 ± 3.0 years), adolescents (N = 11, 17.8 ± 1.1 years), and adults (N = 11, 34.6 ± 11.7 years).

Table 1. Demographic and clinical characteristics of simulated subjects generated using the TIDMS software

Group	N	Age (years)	Weight (kg)	T1D Duration (years)	Total Daily Insulin (U/day)	CR (g CHO/U)	CF (mg/dL per U)	Fasting BG (mg/dL)
Child	11	5.2 ± 3.0	30.5 ± 4.5	2.8 ± 1.7	12.5 ± 3.2	26.5 ± 5.3	117.8 ± 27.7	122.2 ± 5.4
Adolescent	11	17.8 ± 1.1	50.9 ± 8.2	9.7 ± 4.3	35.5 ± 8.9	17.6 ± 7.0	57.1 ± 13.8	120.2 ± 6.0
Adult	11	34.6 ± 11.7	72.9 ± 12.9	17.0 ± 11.3	45.5 ± 10.6	15.9 ± 4.5	42.2 ± 8.0	119.2 ± 4.9

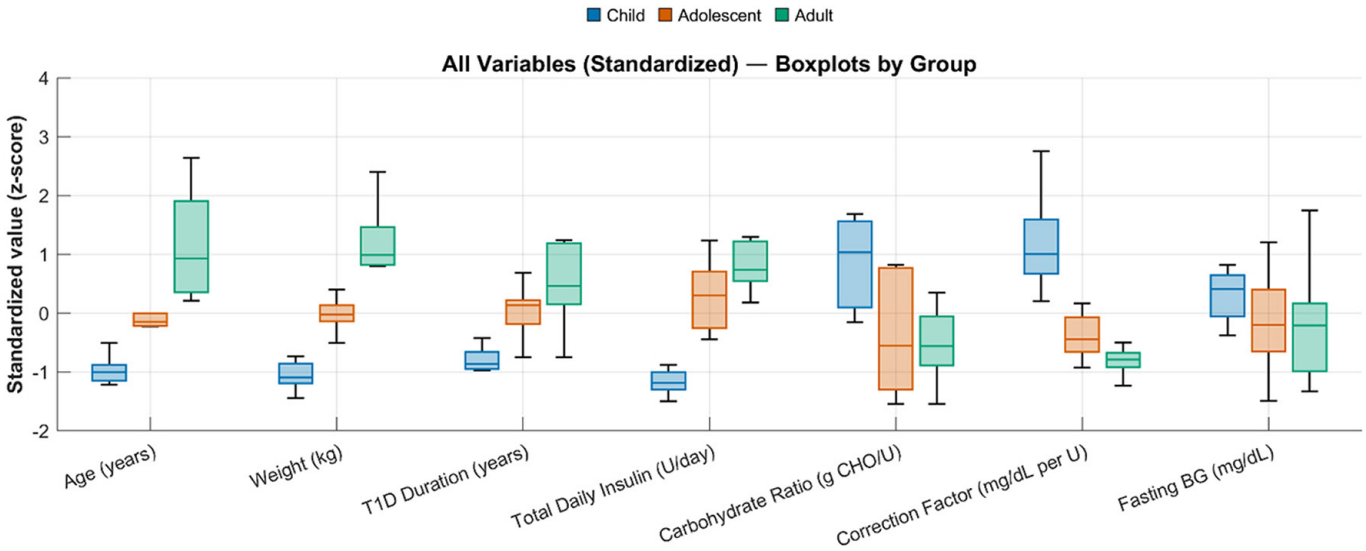

Fig. 1. Standardized (z-score) boxplots of demographic and clinical variables for simulated study subjects generated using the T1DMS software

In T1DMS, each simulation scenario incorporated physiological, hardware, and control parameters representing realistic insulin–glucose dynamics. The CGM sensor

and insulin pump operated at a sampling frequency of 5 minutes, with measurement noise and delivery inaccuracy models enabled to mimic clinical variability. The CGM glucose range was limited to 40–400 mg/dL, and the insulin pump accuracy for bolus delivery was ±5%. For each subject, carbohydrate ratio (CR) and correction factor (CF) were individualized within realistic clinical ranges (15–25 g CHO/U and 40–60 mg/dL per U, respectively), while fasting glucose levels were maintained near 120 mg/dL. Meal quantities and timings were standardized to 70 g (7 a.m.), 100 g (1 p.m.), 30 g (5 p.m.) , and 80 g (8 p.m.). The duration of each simulation was one week under a basal-bolus insulin regimen. Table 1 and Figure 1 show the demographic and clinical characteristics of the simulated subjects and show the normalized parameter outcomes across the different age groups.

Although the UVA/Padova T1DMS provides physiologically consistent and reproducible glucose–insulin dynamics across age groups, it does not capture all sources of variability present in real-world CGM data. In particular, factors such as unreported behavioral deviations, sensor artifacts, acute stress responses, and irregular physical activity are not explicitly modeled. As a result, the in silico findings should be interpreted as characterizing idealized physiological dynamics rather than fully uncontrolled clinical conditions.

2.2 Data preprocessing

The time-series data generated by the T1DMS CGM system underwent a comprehensive preprocessing pipeline to preserve signal integrity and standardize the dataset across all the simulated subjects before model training. Values that were not within the physiologically reasonable range of 40–400 mg/dl were removed as outliers. In addition, missing or irregularly sampled points were imputed using linear interpolation to maintain consistent temporal resolution. A short moving-average filter was applied to attenuate sensor noise while preserving the underlying glucose dynamics. Subsequently, z-scoring of each time series was conducted, which allowed the LSTM models to converge faster and minimize inter-subject variance. The refined glucose signals were segmented into 60-minute overlapping sliding windows; each window was used as an input for one-minute-ahead prediction to capture short-term glucose dynamics and provide higher temporal resolution for interpolation and early trend detection. This design choice is particularly relevant for closed-loop and clinical decision-support systems, where finer temporal forecasts can facilitate smoother insulin delivery adjustments and anticipatory safety mechanisms between CGM updates. The dataset was partitioned into subject-independent training (70%), validation (15%), and testing (15%) subsets. All preprocessing and analytical tasks were performed in MATLAB R2024b (MathWorks, Natick, MA, USA).

2.3 LSTM model architecture

The LSTM network was designed to forecast glucose concentration one minute at (time $t + 1$) based on the proceding 60 minutes of continuous glucose surveillance measurements generated by the T1DMS simulator. The LSTM architecture was selected as it can effectively capture non-linear and time-correlated physiological signals, especially those arising from glucose-insulin feedback mechanisms. Each model was fed with one input feature, the glucose concentration, with a sequence length of 60 samples, corresponding to one hour of historical data.

The network model consisted of a sequence input layer followed by an LSTM layer with 128 hidden units configured in last-output mode to ensure that a long-term temporal structure is maintained. A sequence of fully connected layers of 150, 100, 50, and 20 neurons was fed with the output of the LSTM, each with the Rectified Linear Unit (ReLU) function to introduce nonlinearity. Two dropout layers (0.20 and 0.15) were applied during training to reduce overfitting by randomly deactivating neurons. The final output layer contained a single neuron to produce continuous glucose predictions.

Training was performed in MATLAB R2024b (MathWorks, Natick, MA, USA) using the Deep Learning Toolbox and optimized with the Adam optimizer (learning rate = 0.001, gradient decay factor (β_1) = 0.9, and squared-gradient decay factor (β_2) = 0.999). The mean squared error (MSE) was used as the loss function, and the root mean squared error (RMSE), along with other pertinent statistical metrics, was utilized to evaluate the model performance. The training was conducted over 200 epochs with a batch size of 128 samples, and early stopping was used after the validation loss plateaued without improvement for 10 consecutive epochs. Hyperparameters, including the number of hidden units, dropout rates, and learning rate, were tuned using five-fold cross-validation on the training dataset. Initial weights were initialized using the Glorot uniform method, and the gradients were ensured during optimization.

2.4 Experimental design

Two assessment protocols were implemented in the study. The first protocol evaluated the model performance within a single age group (within-cohort), whereas the second assessed performance across different age groups (cross-cohort). The within-cohort protocol was used to examine glucose patterns specific to each age subgroup, thereby determining the model's ability to detect age-dependent temporal dynamics. The cross-cohort experiment involved training the model on a given age group and testing it on another.

Experiments were conducted using virtual test subjects generated by the T1DMS simulator, which models physiological processes across all ages while accounting for realistic inter-individual variability in insulin sensitivity, carbohydrate-to-insulin ratio, and glucose dynamics. A Leave-One-Subject-Out (LOSO) cross-validation approach was used to evaluate model performance and generalization within each group. This approach produces robust and unbiased estimates of model performance while accounting for individual variability, which is imperative to consider when analyzing physiological recordings.

To examine whether simple adaptation mechanisms can mitigate age-related performance degradation, a transfer-learning–based strategy was applied to selected cross-cohort scenarios. Specifically, pretrained LSTM models were adapted to the target age cohort by fine-tuning only the final output layer using a small subset of target-cohort data, while all recurrent layers were kept fixed. This approach enables efficient adaptation to age-dependent physiological differences without full model retraining.

2.5 Performance evaluation metrics

The predictive performance of the glucose prediction models was evaluated using five standard statistical measures commonly used in glucose prediction research: mean absolute error (MAE), RMSE, normalized root mean squared error

(NRMSE), coefficient of determination (R^2), and Pearson's correlation coefficient (r). In addition, Bland–Altman analysis was employed to assess the model agreement and clinical accuracy.

The MAE and RMSE quantify the mean absolute and mean squared differences between predicted and observed glucose values, respectively, and they are defined as follows:

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

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

where N is the total amount of glucose samples, y_i is the actual glucose value at the ith time step, and $\hat{y}_i$ is the corresponding predicted value [24], [25].

To allow a fair comparison across subjects and age groups with varying glucose ranges, RMSE was normalized as follows:

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \tag{3}$$

where y_{max} and y_{min} denote the maximum and minimum glucose values, respectively [26].

The coefficient of determination (R^2) quantifies the proportion of variance in the reference glucose signal that could be attributed to model predictions and is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

where $\bar{y}$ is the mean value of the actual glucose values [27].

The Pearson's correlation coefficient (r) measures the strength of the linear relationship between the predicted and measured glucose values, representing the time conformity and predictability:

$$r = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{\hat{y}})^2}} \tag{5}$$

where $\bar{\hat{y}}$ is the mean of the predicted glucose values [28].

Bland–Altman analysis, also known as difference plot or Tukey mean-difference plot, was performed to measure the level of agreement and correlation between predicted and actual glucose measurements by computing the estimate of bias and 95% limits of agreement (mean bias ± 1.96 SD of the differences) [29]. The difference between the predicted and actual glucose values was plotted against their averages. Finally, nonparametric statistical tests were applied to compare model performance across cohorts. Multi-group comparisons were conducted using the Kruskal–Wallis test followed by Dunn–Šidák post hoc analysis, while two-group comparisons employed the Wilcoxon rank-sum test. Statistical significance was set at $p < 0.05$ [30], [31].

3 RESULTS

3.1 Glucose dynamics across age groups in T1DMS simulations

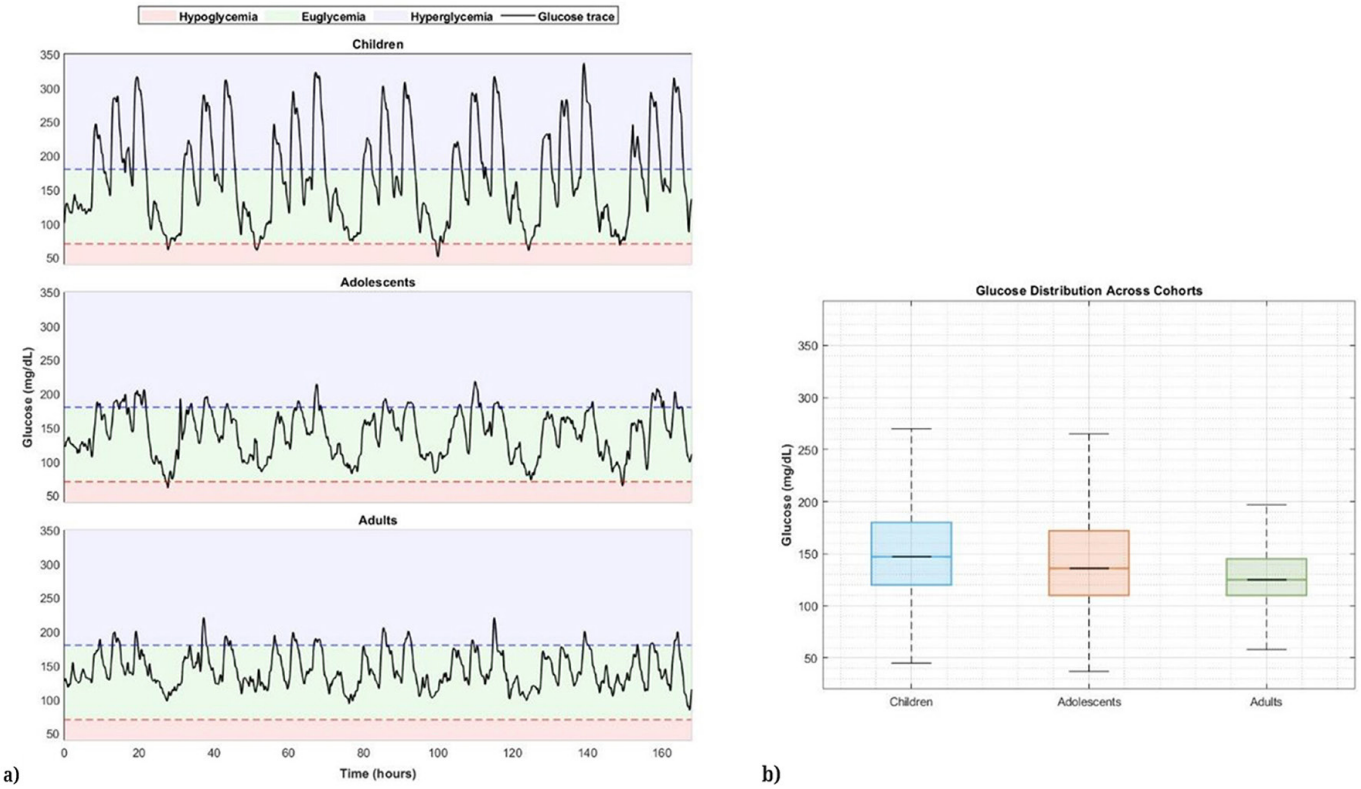

Fig. 2. (a) Representative CGM traces for three cohorts showing glucose dynamics across approximately 7 days (1-minute sampling). (b) Distribution of CGM values across the three cohorts

Figure 2a shows CGM traces of simulated children, adolescents, and adults over a period of seven days, demonstrating the unique glycemic dynamics characteristic of each age group under comparable meals and insulin conditions. Glucose excursions were significantly wider in children, with frequent peaks above 250 mg/dL and troughs below 70 mg/dL, reflecting their higher susceptibility to both hyper- and hypoglycemia. These variations indicate hypersensitivity to insulin and metabolic variability that exists in younger T1D patients. The patterns among adolescent subjects were stabilized to some extent; however, the traces showed typical postprandial peaks and intermittent hypoglycemic dips, which are expected due to physiological and behavioral variations during puberty. In the adult subjects, the oscillations were smaller, and the glucose trajectories became more stable, resulting in fewer boundary crossings of glucose trajectories.

The three background color regions in Figure 2a represent clinically defined glycemic ranges: hypoglycemia (<70 mg/dL) in red, euglycemia (70–180 mg/dL) in green, and hyperglycemia (>180 mg/dL) in blue. The thresholds applied in computing Time-Below-Range (TBR), Time-In-Range (TIR), and Time-Above-Range (TAR), based on the International Consensus on Time-in-Range, are shown as horizontal dashed lines. The trend lines indicate that the adult subjects remained within the normal range compared with the other cohorts. These differences between the glycemic variation of younger and older participants can be interpreted in the context of T1D management systems.

Figure 2b shows the CGM distributions of the three simulated cohorts. The children and the adolescents demonstrated higher medians and interquartile ranges than adults (children: median ≈ 150 mg/dL, IQR ≈ 120–180 mg/dL; adolescents: median ≈ 140 mg/dL, IQR ≈ 110–170 mg/dL), reflecting greater euglycemic variability and anomalies. The glucose range for children extended roughly from 50 mg/dL to 260 mg/dL, thereby causing greater oscillations in Figure 2a. Even though the dispersion was low in adolescents, their upper tails were similar, providing evidence of metabolic changes during puberty. The median glucose was lower in the adult cohort, approximately 120 mg/dL, and the interquartile range was smaller (around 90–150 mg/dL), indicating superior glycemic control in this cohort under the same simulated conditions. The quantitative decrease in central tendency and variability of the glucose concentration with age justifies the observed physiological enhancement of insulin sensitivity and glucose uptake in the T1DMS.

3.2 TIR and glycemic metrics

Table 2. Summary of glycemic control and variability metrics across cohorts (children, adolescents, and adults)

Metric	Children	Adolescents	Adults	Group Comparison	p-Value
Mean glucose (mg/dL)	154.3 ± 11.5	145.1 ± 25.3	129.2 ± 9.0	a a b	0.007
CV (%)	28.4 ± 6.8	28.1 ± 4.6	19.7 ± 2.8	a a b	p < 0.001
TIR (70–180 mg/dL, %)	74.3 ± 12.4	78.0 ± 16.3	94.9 ± 4.0	a a b	p < 0.001
TBR (<70 mg/dL, %)	0.8 ± 1.9	2.2 ± 3.3	0.1 ± 0.3	ns	0.135
TBR (<54 mg/dL, %)	0.1 ± 0.2	0.3 ± 0.7	0.02 ± 0.01	ns	0.187
TAR (>180 mg/dL, %)	24.9 ± 11.4	19.8 ± 17.9	5.0 ± 4.1	a a b	0.001
TAR (>250 mg/dL, %)	4.5 ± 5.6	4.5 ± 7.8	0.0 ± 0.0	a a b	0.007
MAGE (mg/dL)	92.1 ± 30.5	82.0 ± 31.7	52.0 ± 8.8	a a b	0.001
$CONGA_{1h}$ (mg/dL)	40.5 ± 10.1	34.2 ± 9.7	26.2 ± 4.0	a a b	0.005
LBGI	1.5 ± 1.7	2.0 ± 1.7	0.7 ± 0.4	ns	0.069
HBGI	6.7 ± 3.0	5.8 ± 4.2	2.3 ± 0.7	a a b	0.001

Notes: Data are mean ± standard deviation (SD) across subjects. Significance was assessed by the Kruskal–Wallis test with Dunn–Šidák post-hoc comparisons. Different superscript letters (a, b) denote statistically distinct groups ($p < 0.05$); "ns" = non-significant.

Table 2 summarizes the amount of glycemic control and variability indices of the three simulated age groups. Kruskal–Wallis tests were used to statistically compare each of the metrics between the groups, followed by the Dunn–Šidák post-hoc pairwise tests of multiple comparisons. The table uses a different superscript letter to indicate significant differences ($p < 0.05$).

The findings have demonstrated significant age-specific increases in glucose control. Adults exhibited significantly lower mean glucose levels (129.2 ± 9.0 mg/dL) than both children (154.3 ± 11.5 mg/dL) and adolescents (145.1 ± 25.3 mg/dL; $p = 0.007$). The coefficient of variation (CV) similarly declined from approximately 28% in younger cohorts to 19.7 ± 2.8% in adults ($p < 0.001$), reflecting reduced glycemic variability with advancing age. The time-in-range (TIR, 70–180 mg/dL) was highest in adults (94.9 ± 4.0%), representing a significant improvement over children

(74.3 ± 12.4%) and adolescents (78.0 ± 16.3%; $p < 0.001$). Conversely, TAR decreased sharply with age, both at the >180 mg/dL threshold (24.9 ± 11.4% to 5.0 ± 4.1%) and the >250 mg/dL threshold (4.5 ± 5.6%; $p \leq 0.007$), indicating enhanced postprandial control in adults.

Although all groups had minimal time below range (TBR), adolescents had a slightly greater mean TBR than children and adults, which was not statistically significant. Similar trends were seen in glycemic variability. Adults' first-hour mean amplitude of glycemic excursions (MAGE) and continuous overlapping net glycemic action ($CONGA_{1h}$) ratings were considerably lower than those of younger cohorts. The high blood glucose index (HBGI) was substantially lower in adults (2.3 ± 0.7) compared to children (6.7 ± 3.0) and adolescents (5.8 ± 4.2), but the low blood glucose index (LBGI) did not change significantly ($p = 0.069$). The nonparametric analyses show that adults had a lower mean glucose level, less fluctuation, and higher TIR than children and adolescents. The T1DMS shows physiological changes in insulin sensitivity and metabolic control.

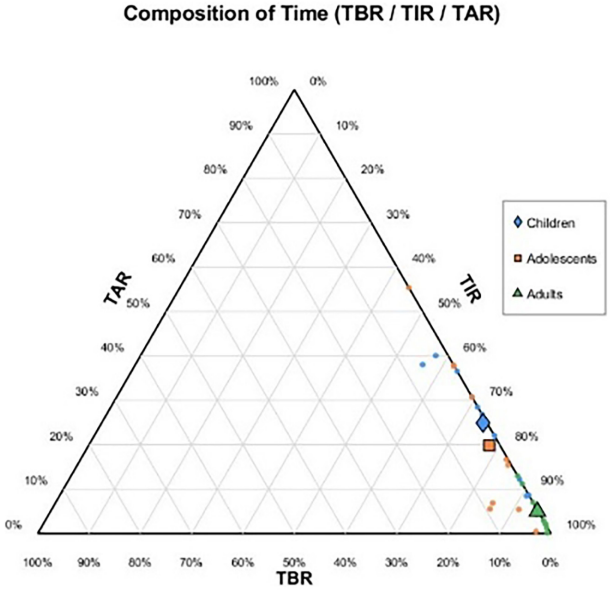

Fig. 3. Ternary plot showing the composition of time spent in three glycemic ranges: Time Below Range (TBR, <70 mg/dL), Time in Range (TIR, 70–180 mg/dL), and Time Above Range (TAR, >180 mg/dL) for each subject across the three cohorts

Figure 3 shows the composition of time spent in TBR (<70 mg/dL), TIR (70–80 mg/dL), and TAR (>180 mg/dL) across the three cohorts. Each point represents an individual simulated subject, with cohort membership indicated by color and marker shape: blue diamonds for children, orange squares for adolescents, and green triangles for adults. The cohort data points cluster near TIR, indicating that most of the simulated subjects are predominantly within the clinically recommended glucose range. Adult subjects are concentrated near the apex of the TIR region, reflecting higher exposure to TIR and minimal exposure to TAR. In contrast, child subjects are distributed along the TAR axis, indicating a high load of hyperglycemia and high inter-subject variability. The quantitative results summarized in Table 2 support this ternary distribution, explaining that glycemic control improves with age. The limited aggregation of adults inside the euglycemic zone corroborates

the T1DMS model's capacity to reproduce age-specific physiological responses under standardized insulin and meal regimens.

3.3 Within- and cross-cohort generalization of LSTM models

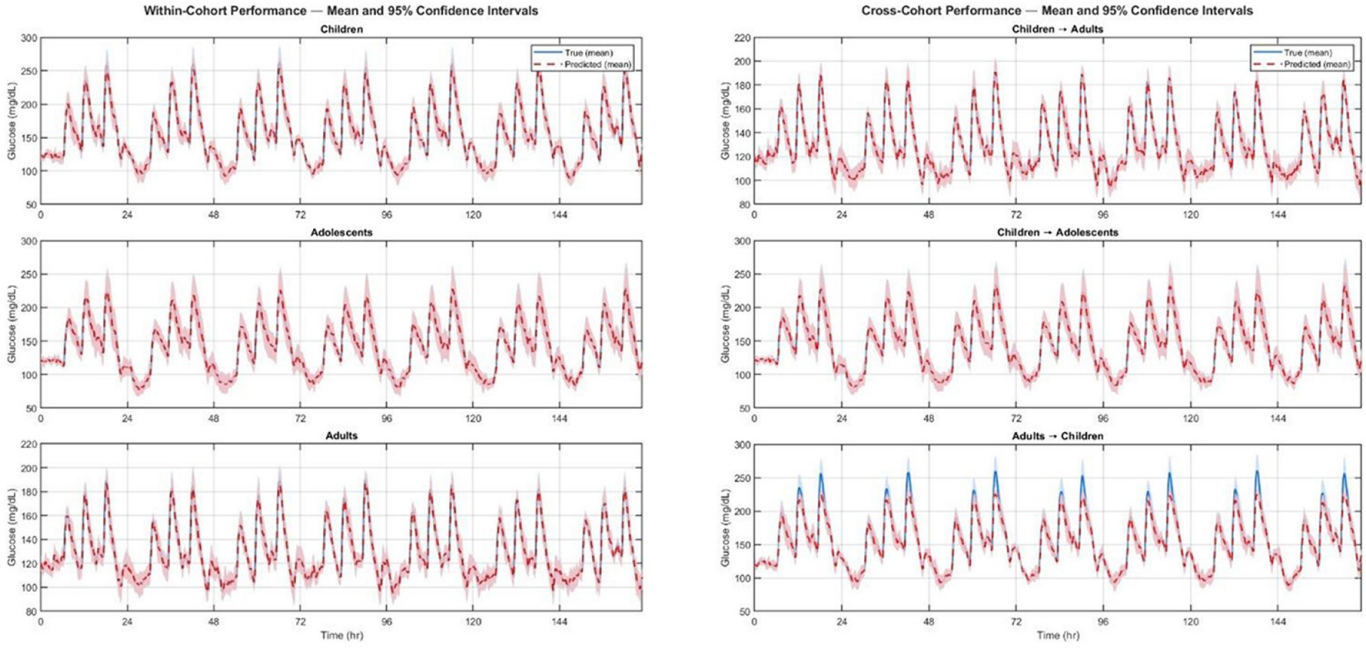

Fig. 4. Mean and 95% confidence intervals of predicted versus true glucose concentrations across test subjects for the three cohorts

Notes: The left column illustrates within-cohort performance, and the right column shows cross-cohort generalization. The shaded regions indicate 95% confidence intervals.

Figure 4 illustrates the within-cohort and cross-cohort validation performance of LSTM-trained and assessed models for children, adolescents, and adults across within- and cross-cohorts. Each panel shows true mean glucose (solid blue) and predicted mean glucose (dashed red). The shaded areas indicate ±95% confidence intervals across subjects during the testing period. The left column presents within-cohort results, and the right column presents cross-cohort generalization; the remaining cross-cohort combinations are provided in the Supplementary Figure S1. The models successfully captured the time-varying characteristics of glucose and insulin in T1DMS, as evidenced by the strong agreement between the predicted and observed glucose curves across all cohorts. The substantial overlap between the observed and predicted mean profiles indicates high short-term predictive accuracy and minimal bias, while the span of the confidence intervals suggests that the model remained stable across subjects within each cohort, with only slight deviations during postprandial glucose spikes.

Adults exhibited the narrowest range of prediction envelopes due to their relatively low metabolic variability, while children showed wider confidence ranges, especially around meals. Physiological variations in adolescents resulted in intermediate results accompanied by precise yet more variable predictions. Overall, cohort-specific LSTM models demonstrated the ability to generalize to new subjects within their respective age groups and properly reproduced the simulator glucose dynamics, establishing a robust basis for cross-cohort investigations.

3.4 Within- and cross-cohort performance evaluation

Table 3. Intra- and cross-cohort evaluation results of LSTM-based glucose prediction models across age groups, with selected transfer-learning–based fine-tuning of the final output layer (Adapted models are highlighted for clarity)

Training	Testing	MAE (mg/dL)	RMSE (mg/dL)	NRMSE	Pearson's r	R^2
Children	Children	1.54 ± 0.68	2.46 ± 1.64	0.01 ± 0.005	0.999 ± 0.001	0.997 ± 0.002
	Adults	0.99 ± 0.09	1.21 ± 0.09	0.05 ± 0.008	0.999 ± 0.0001	0.998 ± 0.001
	Adolescents	2.43 ± 1.84	4.30 ± 4.12	0.09 ± 0.052	0.998 ± 0.003	0.989 ± 0.013
Children (fine-tuned on adolescents)	**Adolescents**	**2.02 ± 0.91**	**3.14 ± 2.02**	**0.05 ± 0.001**	**0.998 ± 0.0002**	**0.989 ± 0.003**
Adolescents	Adolescents	0.95 ± 0.32	1.35 ± 0.94	0.007 ± 0.002	0.999 ± 0.0004	0.998 ± 0.001
	Adults	1.61 ± 0.05	1.76 ± 0.05	0.07 ± 0.01	0.999 ± 0.0001	0.995 ± 0.002
	Children	0.98 ± 0.19	1.27 ± 0.31	0.03 ± 0.004	0.999 ± 0.0001	0.999 ± 0.0003
Adults	Adults	0.85 ± 0.15	1.19 ± 0.42	0.008 ± 0.002	0.999 ± 0.0003	0.997 ± 0.001
	Adolescents	4.05 ± 4.89	7.97 ± 10.19	0.15 ± 0.15	0.988 ± 0.018	0.957 ± 0.070
	Children	5.74 ± 2.76	9.41 ± 5.70	0.20 ± 0.06	0.992 ± 0.006	0.958 ± 0.026
Adults (fine-tuned on children)	**Children**	**3.03 ± 1.74**	**6.99 ± 4.61**	**0.14 ± 0.08**	**0.997 ± 0.009**	**0.960 ± 0.019**
Adults (fine-tuned on adolescents)	**Adolescents**	**3.02 ± 1.55**	**4.80 ± 2.83**	**0.12 ± 0.05**	**0.992 ± 0.009**	**0.972 ± 0.044**

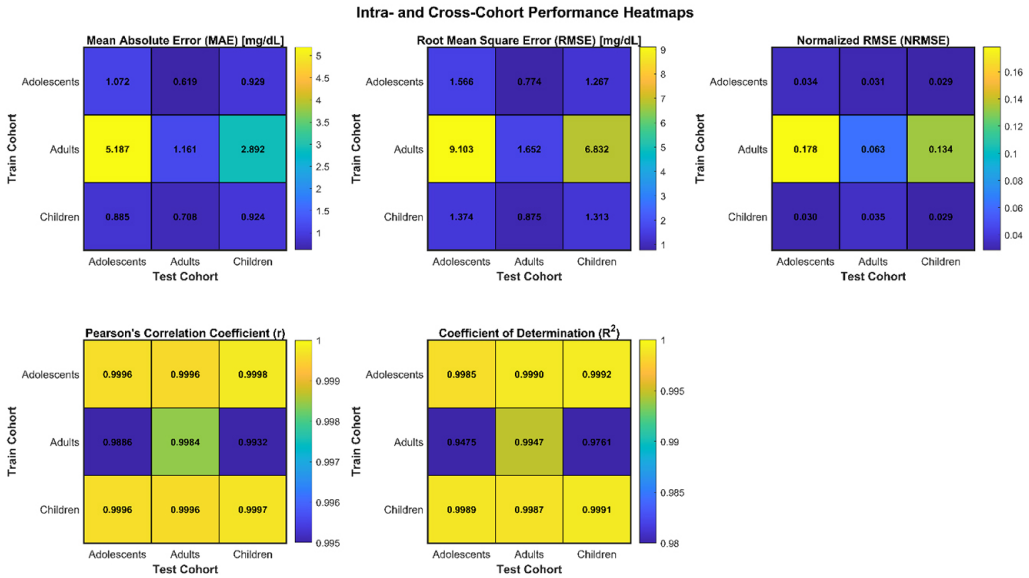

Fig. 5. Comprehensive intra- and cross-cohort performance heatmaps for glucose prediction models

Each LSTM model, trained separately on children, adolescents, and adults, was evaluated on all three age cohorts to assess their generalization capability. The results are shown in Table 3 and Figure 5. The diagonal entries of the matrices represent within-cohort evaluations. These within-group models achieved the best performance across all metrics. In addition, they exhibited nearly perfect correlations ($r \approx 0.999$, $R^2 \approx 0.997$) and extremely low prediction errors (MAE ≈ 1–2 mg/dL, RMSE ≈ 1–2.5 mg/dL, NRMSE < 0.01). This confirms the models' strong ability to predict the glucose–insulin dynamics of each age group.

In cross-cohort assessments, models showed strong generalization asymmetries. Models trained on adult data exhibited the greatest performance degradation when

applied to children (MAE = 5.74 ± 2.76 mg/dL, NRMSE = 0.20 ± 0.06) and adolescents (MAE = 4.05 ± 4.89 mg/dL, NRMSE = 0.15 ± 0.15). Physiological differences such as insulin sensitivity, metabolic rate, and glycemic variability limit the applicability of adult-derived models to juvenile populations, explaining the decrease in prediction accuracy. Conversely, models trained on data from children or adolescents generalized more effectively to the adult cohort.

In cross-cohort evaluations exhibiting pronounced performance degradation, the proposed lightweight adaptation strategy led to consistent error reductions, as shown in Table 3. In particular, adapting adult-trained models to pediatric and adolescent cohorts substantially reduced prediction errors compared to non-adapted baselines. These results demonstrate that even minimal output-layer adaptation can partially compensate for age-related physiological mismatch and improve cross-age prediction accuracy.

As shown in Figure 5, the significant variance in the diagonal and off-diagonal segments of the matrix demonstrates the physiological influence of age on prediction accuracy. The findings indicate that LSTM networks excel in intra-cohort glucose dynamics but are constrained by intergroup physiological variability. To obtain a robust, ageless, and unified glucose forecasting output, successive models may require cohort-specific training or adaptive transfer learning.

3.5 Bland–Altman agreement analysis

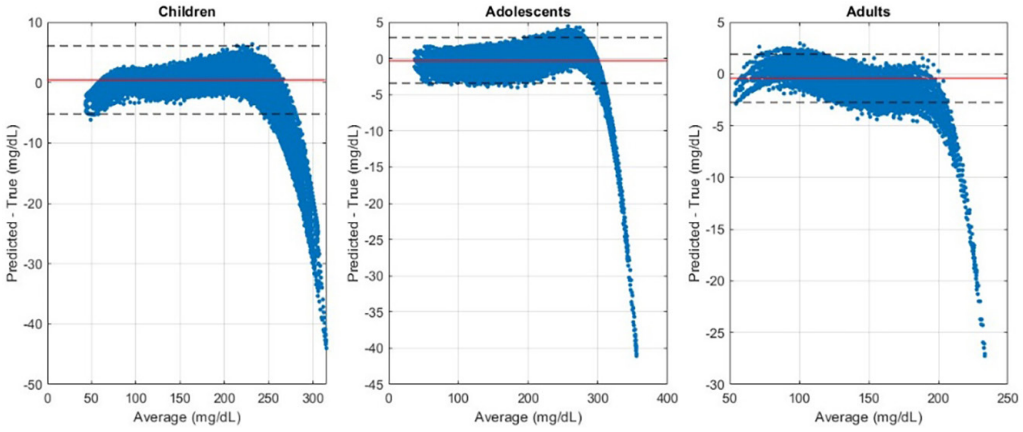

Fig. 6. Bland–Altman plots illustrating within-cohort agreement between predicted and true glucose concentrations for children, adolescents, and adults

Note: The central red line represents the mean bias, while dashed lines denote the 95% limits.

Figure 6 presents the Bland–Altman plots comparing predicted and measured glucose concentrations for children, adolescents, and adults during within-cohort evaluation. Across all age groups, data points are densely distributed around the zero-bias line, and the limits of agreement remain relatively narrow, indicating high concordance between model predictions and reference values.

For every cohort, the mean bias was close to zero (within ±5 mg/dL), confirming that the LSTM models did not systematically over- or underestimate glucose levels. Slightly broader dispersion was observed at higher glucose concentrations (>250 mg/dL), primarily reflecting the nonlinear physiological behavior of glucose–insulin interactions during hyperglycemia rather than model error. Consistent with lower glycemic variability, the adult cohort exhibited the tightest limits of agreement,

whereas children showed wider limits due to larger postprandial excursions. These findings support the reliability and clinical acceptability of the within-cohort LSTM models, showing that the errors accounted by the predictive model are symmetrically distributed around the physiological glucose variability.

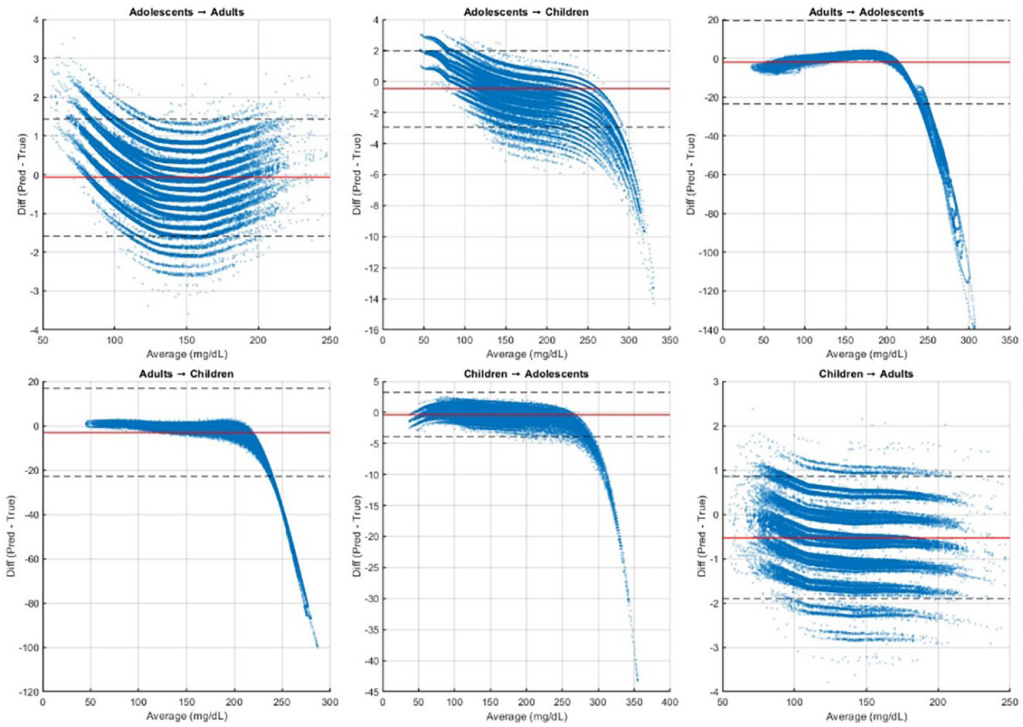

Fig. 7. Bland–Altman plots illustrating cross-cohort agreement between predicted and true glucose concentrations for children, adolescents, and adults

Notes: The central red line represents the mean bias, while dashed lines denote the 95% limits of agreement. The solid red line denotes the mean bias between predicted and reference glucose values, while the dashed lines mark the 95 % limits of agreement (bias ± 1.96 × SD).

Figure 7 shows the Bland–Altman analysis to evaluate cross-cohort comparisons, illustrating the concordance between predicted and true glucose concentrations across age groups. Models trained on children and adolescent cohorts exhibited limited agreement and moderate bias when applied to other cohorts, suggesting more resilience and flexibility in these age groups. The adolescent-to-adult and child-to-adult comparisons exhibited similar bias magnitudes (within ±5 mg/dL) with uniform distribution across the glucose spectrum, signifying inter-cohort physiological variability. In contrast, adult-trained models tended to overestimate hyperglycemia when applied to children and adolescent subjects. This result agrees with the performance asymmetries observed in the heatmaps (see Figure 5) and confirms that models trained on younger cohorts achieved superior generalization.

4 DISCUSSION

In this study, the FDA-approved UVA/Padova T1D Metabolic Simulator (T1DMS) was used to evaluate the overall performance of LSTM networks in predicting short-term glucose dynamics across different age cohorts. The results indicate that it is challenging to develop prediction models capable of maintaining stable performance across diverse physiological conditions.

Age exerts a measurable effect on glycemic interactions, as indicated by the simulated glucose traces. Children exhibited high intrinsic variability and rapid glucose kinetics. These trends correspond to physiological factors such as increased insulin sensitivity, irregular carbohydrate consumption, and energy loss as a result of physical exercise. In adolescents, hormonal variation and behavioral conditions during puberty induced momentary postprandial peaks and episodes of hypoglycemia [32]. In contrast, adults have more regulated glucose profiles due to predictable insulin intake, reduced glucose transport, and more consistent dietary and dosing routines [33], [34]. The age-stratified glucose modelling structure of the T1DMS effectively regenerates these patterns, consistent with findings described in clinical reports [35].

Adults achieved the highest TIR and the lowest TAR, along with smaller variability indices such as the MAGE and continuous net glycemic assessment (CONGA). On the other hand, younger age groups showed a high rate of readings that were not within the stipulated target range and higher fluctuation indices, which are indicative of strong glycemic variability in children and adolescents. This is consistent with reported daily glucose variation in children and adolescent T1D patients under the care of health professionals and utilizing more intensive and flexible insulin plans. The cohort-specific LSTM models were exemplary predictors when trained and tested on data from matched cohorts. The predicted glucose trajectories followed the reference signals with minimal bias and near-unity correlation coefficients. Such short-term dependency models are very realistic simulations of the key physiological feedback loops that control blood glucose, such as insulin absorption, hepatic gluconeogenesis, and meal-induced perturbations. The theoretical optima were in very close alignment with the empirical results, which means the trained networks captured temporal and metabolic characteristics unique to each cohort.

Conversely, the cross-cohort evaluations showed very strong performance differences, reflecting physiological differences in glucose variability. Models trained on adult data showed the greatest decline in accuracy when applied to children and adolescent cohorts, as indicated by increased MAE, RMSE, and reduced Bland–Altman concordance. The limited glycemic variability and dampened metabolic dynamics in adults might limit the transferability of glucose-variability relationships to younger age groups whose dynamics are faster and more vigorous. On the other hand, models trained on children and adolescent data generalized better to adult cohorts. The broader range of glycemic oscillations and metabolic responses in the younger groups provides a more diverse feature space, allowing the network to relate temporal input sequences to glucose outputs. These results indicate that the predictive accuracy depends on the network architecture and representativeness of the physiological data.

From a modeling perspective, the observed age-related degradation in cross-cohort performance can be interpreted as a form of domain uncertainty that primarily affects the mapping between learned temporal patterns and glucose predictions. To address this, a lightweight transfer-learning–based adaptation strategy was introduced by fine-tuning only the final output layer of pretrained LSTM models while keeping all recurrent layers fixed. This limited adaptation preserves the core temporal dynamics learned by the network while compensating for systematic age-dependent shifts in glucose–insulin behavior. The resulting performance improvements in challenging cross-age scenarios demonstrate that simple adaptation mechanisms can effectively enhance robustness without requiring full model retraining or architectural modification.

The findings of this work have implications for the clinical treatment of diabetes. Children with rapid and large glycemic swings require accurate short-term glucose

modeling to reduce cases of acute glycemic oscillation and improve insulin delivery. There is a strong requirement for careful calibration of closed-loop insulin delivery systems and clinical decision-support tools since model behavior is sensitive to physiological factors. CGM devices can improve their safety and effectiveness by incorporating adaptive prediction algorithms that personalize glucose forecasts. Simulation-based research is a cost-effective and ethically sound method for assessing the performance of various algorithms across a wide range of physiological conditions before conducting clinical trials.

Although the present study focuses on open-loop glucose prediction, the proposed age-aware modeling framework has direct implications for closed-loop insulin delivery and clinical decision-support systems. In closed-loop settings, prediction accuracy is a critical factor influencing insulin dosing decisions, safety margins, and the avoidance of hypoglycemic events. Improved cross-age prediction performance, particularly in pediatric and adolescent populations, can support the development of age-adaptive controllers that dynamically adjust insulin delivery based on patient-specific physiological characteristics. Similarly, in clinical decision-support applications, reliable short-term glucose forecasts can enhance alarm systems, dosing recommendations, and clinician oversight without directly controlling insulin infusion. Future work will investigate the integration of the proposed prediction models into closed-loop control architectures to quantitatively assess their impact on insulin delivery outcomes and safety.

5 CONCLUSIONS

The FDA-approved UVA/Padova T1D Metabolic (T1DMS) was used to evaluate the glucose prediction models based on LSTM networks across different age groups. The cohort-specific models achieved better accuracy compared to the models trained on aggregated age groups, demonstrating the importance of age-dependent physiological dynamics of glucose–insulin regulation in children, adolescents, and adults. Models trained on younger cohorts demonstrated greater adaptability when applied to adult populations due to the metabolic variability inherent in these age groups. Simplified simulations were used to illustrate model behavior under controlled conditions. However, given the in silico nature of this study, real-world confounding variables such as stress, circadian rhythms, and physical activity were not incorporated. The findings of this study can be validated by future research using clinical datasets and state-of-the-art physiological modeling techniques to develop and examine adaptive, physiology-based glucose prediction algorithms that provide high-quality predictions despite physiological fluctuations.

6 REFERENCES

[1] NCD Risk Factor Collaboration (NCD-RisC), "Worldwide trends in diabetes since 1980: A pooled analysis of 751 population-based studies with 4.4 million participants," *The Lancet*, vol. 387, pp. 1513–1530, 2016.

[2] Emerging Risk Factors Collaboration *et al.*, "Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: A collaborative meta-analysis of 102 prospective studies," *The Lancet*, vol. 375, pp. 2215–2222, 2010. https://doi.org/10.1016/S0140-6736(10)60484-9

[3] J. M. Forbes and M. E. Cooper, "Mechanisms of diabetic complications," *Physiol. Rev.*, vol. 93, pp. 137–188, 2013. https://doi.org/10.1152/physrev.00045.2011

[4] D. M. Nathan, "Long-term complications of diabetes mellitus," *N. Engl. J. Med.*, vol. 328, no. 23, pp. 1676–1685, 1993. https://doi.org/10.1056/NEJM199306103282306

[5] L. Y. Melendez-Ramirez, R. J. Richards, and W. T. Cefalu, "Complications of type 1 diabetes," *Endocrinology and Metabolism Clinics of North America*, vol. 39, pp. 625–640, 2010. https://doi.org/10.1016/j.ecl.2010.05.009

[6] R. J. McCrimmon and R. S. Sherwin, "Hypoglycemia in type 1 diabetes," *Diabetes*, vol. 59, pp. 2333–2339, 2010. https://doi.org/10.2337/db10-0103

[7] M. S. Alzboon, M. S. Al-Batah, M. Alqaraleh, A. Abuashour, and A. F. Hamadah Bader, "Early diagnosis of diabetes: A comparison of machine learning methods," *Int. J. Online Biomed. Eng.*, vol. 19, no. 15, pp. 144–165, 2023. https://doi.org/10.3991/ijoe.v19i15.42417

[8] N. R. Dzakiyullah, M. A. Burhanuddin, R. R. Raja Ikram, N. Yudistira, M. R. Fauzi, and D. J. Purbohadi, "Multi-label risk prediction of diabetes complications using machine learning models," *Int. J. Online Biomed. Eng.*, vol. 20, no. 16, pp. 66–88, 2024. https://doi.org/10.3991/ijoe.v20i16.51643

[9] S. K. Vashist, "Continuous glucose monitoring systems: A review," *Diagnostics*, vol. 3, pp. 385–412, 2013. https://doi.org/10.3390/diagnostics3040385

[10] F. Ståhl and R. Johansson, "Diabetes mellitus modeling and short-term prediction based on blood glucose measurements," *Math. Biosci.*, vol. 217, no. 2, pp. 101–117, 2009. https://doi.org/10.1016/j.mbs.2008.10.008

[11] M. H. Kroll, "Biological variation of glucose and insulin includes a deterministic chaotic component," *Biosystems*, vol. 50, no. 3, pp. 189–201, 1999. https://doi.org/10.1016/S0303-2647(99)00007-6

[12] M. I. Rusydi, R. Kurnia, N. Windasari, and R. Z. Putra, "A stratified modeling–machine learning approach to improve the accuracy of non-invasive blood glucose estimation using photoplethysmography signals," *Int. J. Online Biomed. Eng.*, vol. 21, no. 6, pp. 76–96, 2025. https://doi.org/10.3991/ijoe.v21i06.53815

[13] R. Birjais, A. K. Mourya, R. Chauhan, and H. Kaur, "Prediction and diagnosis of future diabetes risk: A machine learning approach," *SN Applied Sciences*, vol. 1, no. 9, Art. no. 1112, 2019. https://doi.org/10.1007/s42452-019-1117-9

[14] L. Kopitar, P. Kocbek, L. Cilar, A. Sheikh, and G. Stiglic, "Early detection of type 2 diabetes mellitus using machine learning-based prediction models," *Scientific Reports*, vol. 10, no. 1, p. 11981, 2020. https://doi.org/10.1038/s41598-020-68771-z

[15] E. Dritsas and M. Trigka, "Data-driven machine-learning methods for diabetes risk prediction," *Sensors*, vol. 22, no. 14, p. 5304, 2022. https://doi.org/10.3390/s22145304

[16] V. Jaiswal, A. Negi, and T. Pal, "A review on current advances in machine learning based diabetes prediction," *Prim. Care Diabetes*, vol. 15, pp. 435–443, 2021. https://doi.org/10.1016/j.pcd.2021.02.005

[17] M. Rahman, D. Islam, R. J. Mukti, and I. Saha, "A deep learning approach based on convolutional LSTM for detecting diabetes," *Comput. Biol. Chem.*, vol. 88, p. 107329, 2020. https://doi.org/10.1016/j.compbiolchem.2020.107329

[18] S. A. Alex *et al.*, "Deep LSTM model for diabetes prediction with class balancing by SMOTE," *Electronics*, vol. 11, p. 2737, 2022. https://doi.org/10.3390/electronics11172737

[19] S. Soltanizadeh and S. S. Naghibi, "Hybrid CNN–LSTM for predicting diabetes: A review," *Current Diabetes Reviews*, vol. 20, pp. 77–84, 2024. https://doi.org/10.2174/0115733998261151230925062430

[20] P. Dharanyadevi *et al.*, "Comparative analysis of SVM and LSTM models for predicting diabetes: Evaluating performance and interpretability," in *Integrated Technologies in Electrical, Electronics and Biotechnology Engineering*, Boca Raton, FL, USA: CRC Press, 2025, pp. 294–299. https://doi.org/10.1201/9781003606208-50

[21] F. Mohsen *et al.*, "A scoping review of artificial intelligence-based methods for diabetes risk prediction," *NPJ Digit. Med.*, vol. 6, no. 1, p. 197, 2023. https://doi.org/10.1038/s41746-023-00933-5

[22] B. P. Kovatchev, M. Breton, C. Dalla Man, and C. Cobelli, "In silico preclinical trials: A proof of concept in closed-loop control of type 1 diabetes," *J. Diabetes Sci. Technol.*, vol. 3, no. 1, pp. 44–55, 2009. https://doi.org/10.1177/193229680900300106

[23] The Epsilon Group, "Diabetes simulation," 2025. [Online]. Available: https://tegvirginia.com/services/diabetes-simulation/ [Accessed: Oct. 15, 2025].

[24] W. Wang and Y. Lu, "Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 324, p. 012049, 2018. https://doi.org/10.1088/1757-899X/324/1/012049

[25] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not," *Geosci. Model Dev.*, vol. 15, pp. 5481–5487, 2022. https://doi.org/10.5194/gmd-15-5481-2022

[26] J. Freiburghaus, A. Rizzotti, and F. Albertetti, "A deep learning approach for blood glucose prediction of type 1 diabetes," in *Proc. 5th Int. Workshop on Knowledge Discovery in Healthcare Data (KDH)*, 2020, pp. 29–30.

[27] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*. New York, NY, USA: Springer, 2013. https://doi.org/10.1007/978-1-4614-7138-7

[28] E. T. Chen, J. H. Nichols, S. H. Duh, and G. Hortin, "Performance evaluation of blood glucose monitoring devices," *Diabetes Technology & Therapeutics*, vol. 5, pp. 749–768, 2003. https://doi.org/10.1089/152091503322526969

[29] D. Giavarina, "Understanding Bland–Altman analysis," *Biochem. Med.*, vol. 25, pp. 141–151, 2015. https://doi.org/10.11613/BM.2015.015

[30] M. M. M. Mansour, "Non-parametric statistical test for testing exponentiality with applications in medical research," *Stat. Methods Med. Res.*, vol. 29, pp. 413–420, 2020. https://doi.org/10.1177/0962280218824979

[31] S. Thamilselvan and N. A. Ahad, "Comparison of post hoc tests on non-communicable diseases data," *Quantum J. Eng. Sci. Technol.*, vol. 4, p. 12, 2023.

[32] R. A. Williams *et al.*, "Predictors of postprandial glycaemia, insulinaemia and insulin resistance in adolescents," *Br. J. Nutr.*, vol. 125, pp. 1101–1110, 2021. https://doi.org/10.1017/S0007114520003505

[33] D. Montt-Blanchard, R. Sánchez, K. Dubois-Camacho, J. Leppe, and M. T. Onetto, "Hypoglycemia and glycemic variability of people with type 1 diabetes with lower and higher physical activity loads in free-living conditions using continuous subcutaneous insulin infusion with predictive low-glucose suspend system," *BMJ Open Diabetes Research & Care*, vol. 11, p. e003082, 2023. https://doi.org/10.1136/bmjdrc-2022-003082

[34] P. Tsarkova, A. Shestakova, E. Isakov, A. Ibragimova, and M. Shamkhalova, "CSII is related to more stable glycemia in adults with type 1 diabetes," *Endocrine*, vol. 75, pp. 776–780, 2022. https://doi.org/10.1007/s12020-021-02913-9

[35] A. Khadilkar and C. Oza, "Glycaemic control in youth and young adults: Challenges and solutions," *Diabetes Metab. Syndr. Obes.*, vol. 15, pp. 121–129, 2022. https://doi.org/10.2147/DMSO.S304347

7 SUPPLEMENTARY MATERIAL

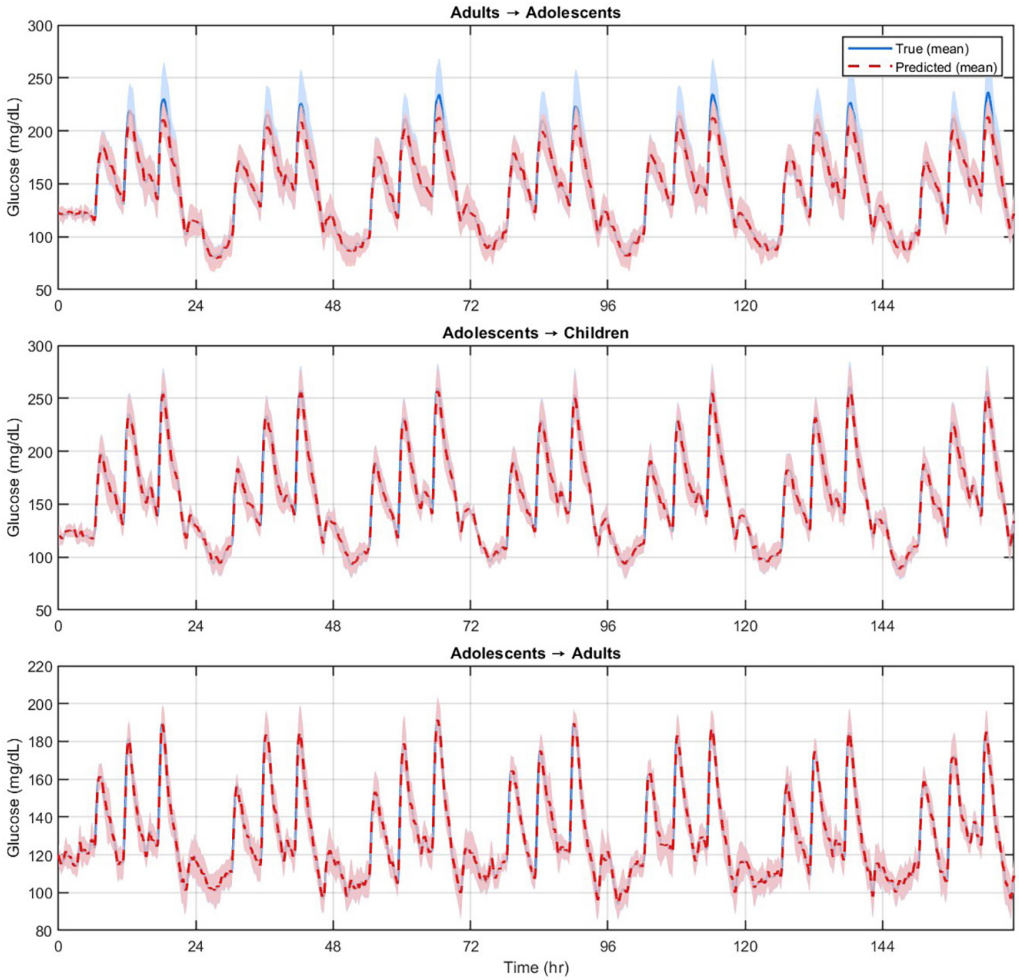

Fig. S1. Cross-cohort LSTM prediction results showing mean true and predicted glucose profiles when models are trained on one age group and tested on another

8 AUTHOR

Saleh I. Alzahrani is an Assistant Professor in the Department of Biomedical Engineering at the College of Engineering, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia. He received his Ph.D. degree in Biomedical Engineering from Colorado State University, Fort Collins, CO, USA. His research interests include machine learning and statistical signal processing for biomedical applications, with a focus on electroencephalography (EEG) analysis and noninvasive brain–computer interfaces (BCIs) (E-mail: sialzahrani@iau.edu.sa).