Sciences Methods and Technologies International Journal (SciMeTech)

ISSN: 3085-5284, journal homepage: www.scimetech.com

TD3-LSTM with Proactive Safety Layer for Closed-Loop Blood Glucose Regulation in Type 1 Diabetes

SAYDTAHIRI Assia[1], ELMAJIDI Azzedine[2], HATIM Anas[1], ELLOUAD Hicham[3]

[1] *LARTID, National School of Applied Sciences of Marrakech, Cadi Ayyad University, Marrakech, Morocco*

[2] *LAREFERENCE, Faculty of Legal Economic and Social Sciences of Marrakech, Cadi Ayyad University, Marrakech, Morocco*

[3] *LEBPBAI, Faculty of medicine and pharmacy of Marrakech, Cadi Ayyad University, Marrakech, Morocco*

How to cite:

SAYDTAHIRI Assia, ELMAJIDI Azzedine, HATIM Anas, ELLOUAD Hicham, "TD3-LSTM with Proactive Safety Layer for Closed-Loop Blood Glucose Regulation in Type 1 Diabetes", Sciences Methods and Technologies International Journal (SciMeTech), (2026) Vol 2, Issue 2, p 69-76

[1] a.saydtahiri.ced@uca.ac.ma [2] elmajidi@gmail.com, [1] hatim.anas@gmail.com [4] h.ellouad.ced@uca.ac.ma

Keywords:

Type 1 Diabetes
Reinforcement Learning
TD3- LSTM
Safety Layer
Artificial Pancreas

Abstract

T1DM demands constant insulin administration to ensure glucose stays within the target window of 70–180 mg/dL. Current deep RL controllers utilizing conventional TD3 possess two major drawbacks — the inability to store past data hinders the recording of the glucose pattern after meals, and the absence of a safety layer leads to a high incidence of hypoglycemia. This study presents TD3-LSTM+PSL, a closed-loop glucose controller featuring an LSTM layer embedded into the TD3 actor-critic framework to incorporate a 2-hour history of CGM readings, complemented by a Proactive Safety Layer (PSL) that estimates future glucose and reduces insulin doses when hypoglycemia risk is detected. The model is evaluated using a simulation study on 10 adult virtual patients utilizing the FDA-approved UVA/Padova T1D simulator without any meal announcements. It demonstrates a significant increase in time-in-range (+10.7%) compared to the standard TD3 baseline (73.7% ± 5.5% vs. 63.0%), together with a reduction in hypoglycemia rate (−10.7%) from 37.0% to 26.3%, and a complete elimination of hyperglycemic episodes (0.0%). These results demonstrate that temporal glucose memory combined with proactive safety intervention substantially improves glycemic control in fully automated closed-loop insulin delivery.

I. Introduction

T1D is an autoimmune disease marked by the breakdown of pancreatic beta cells, leading to the complete lack of insulin and impacting around 8.4 million people across the globe [1]. While in type 2 diabetes, the progression occurs through insulin resistance, type 1 diabetes requires the lifelong supply of exogenous insulin to keep the blood glucose level between the required range of 70-180 mg/dL. The slightest change from this range poses a life threat, as hypoglycemia (below 70 mg/dL) may lead to coma or even sudden death, whereas prolonged hyperglycemia (above 180 mg/dL) exacerbates the cardiovascular and neurological disorders [2]. Modern breakthroughs in signal processing and machine learning have facilitated automatic detection of cardiac problems in patients suffering from diabetes like ventricular ectopic beats [3], thus paving the way for the development of advanced automated management systems.

Nevertheless, the artificial pancreas (AP), which can also be called a closed-loop system for delivering insulin, is the primary technological solution to this issue. By combining a CGM, an insulin pump, and a controller, the devices attempt to imitate the regulatory processes performed by beta cells in healthy subjects [4]. Conventional approaches based on control algorithms such as PID and MPC utilize mathematical models for the glucose-insulin process that need individual calibration to accommodate each case [5]. Neural network-based control schemes have also been investigated as an alternative to the explicit use of physiological models and shown promise for regulating glucose levels in closed-loop systems [6].

DRL has gradually become known as another paradigm that makes it possible to derive the control strategy only based on patient interactions without relying on physiological models explicitly [7]. From amongst recently developed DRL

controllers, G2P2C [4] is currently considered an innovative benchmark due to its use of the Proximal Policy Optimization (PPO) algorithm. However, there exist two inherent structural shortcomings that limit their applicability in a clinical setting. Firstly, being a method dependent on on-policy updates, PPO loses the trajectories generated after each gradient descent, thereby resulting in inefficient sampling. Secondly, most other competing algorithms lack a prospective safety measure whereby the agent can generate locally optimal insulin dosages that may still cause hypoglycemia under unforeseen physiological states.

These issues are both tackled in our study via a joint framework termed TD3-LSTM. Specifically, on the algorithmic side, we substitute the PPO-based controller with a Twin Delayed Deep Deterministic Policy Gradient (TD3) framework [8] that leverages replay buffers for collecting historical transition data and is native to action spaces. On the temporal side, the temporal dynamics needed to represent postprandial changes and delayed effect of insulin is captured using bidirectional LSTM applied to a two-hour glucose history, which is possible without employing an additional glucose prediction sub-model. On the safety side, we present a proactive safety layer (PSL) that intervenes at each candidate's dose suggestion and evaluates the short-term trajectory, attenuating the dose incrementally once hypoglycemia is predicted. This approach can be seen as an extension of the double-safety idea suggested by Zhao et al. [9] to a TD3 setting, thus providing a hard safety constraint operating independently from the trained policy.

The effectiveness of the proposed algorithm was evaluated in silico via an FDA-approved UVA/Padova simulator for type 1 diabetes for 10 adults, who were asked to consume 175 g of carbohydrates daily without any prior notification about meals. As can be seen from the experimental findings, the proposed algorithm is capable of maintaining a high time-in-range (TIR) while lowering hypoglycemic exposure effectively.

This paper is structured in the following way. Section II provides an overview of research related to closed-loop glycemic control. Section III introduces the TD3-LSTM architecture and the proactive safety mechanism. Section IV outlines the experimental procedure. Section V offers the results of the experiment. Section VI identifies limitations and suggests further work. Section VII concludes.

II. Methods

Closed-loop glucose regulation is cast as a POMDP with the tuple (S, A, T, R, O, Ω, γ) [10]. The physiological state vectors belonging to the set S is characterized by the glucose-insulin interaction model in UVA/Padova, which cannot be observed [11]. The observable CGM signal $o \in O$ is affected by Gaussian measurement noise according to the Dexcom G6 sensor specifications [12].

Action space $A = [0, a_{max}]$ is a continuum and corresponds to basal insulin dose in units per minute. Maximum dose, a_{max}, depends on the age cohort: it equals 0.05 u/min for adults. Discount factor is equal to 0.99 and aims at promoting long-term glycemic regulation.

The reward signal is designed asymmetrically to reflect the clinical urgency of hypoglycemia. This penalty term depends on the Magni risk index, which considers deviation from the target zone in an asymmetric manner [13]. While a hypoglycemic event is punished four times harsher than a hyperglycemic one because of its high instantaneous danger, the reward for time t is expressed as:

$$\begin{cases} r(t) = +1.0 \text{ if } BG(t) \in [70, 180]; \\ r(t) = -4 \cdot \varrho(BG(t)) \text{ if } BG(t) < 70; \\ r(t) = -\varrho(BG(t)) \text{ if } BG(t) > 180 \end{cases} \quad (1)$$

where $\varrho(BG) = 10 \cdot [1.509 \cdot (\ln(BG)^{1.084} - 5.381)]^2$ is the Magni risk function. The reward is normalized online using a running mean and standard deviation to stabilize training.

A representation of the complete architecture of the proposed system is shown in Figure 1. This model comprises four components: the environment, LSTM encoder to learn glucose temporal history, TD3 actor-critic controller, and PSL – a hard safety constraint imposed before insulin injection.

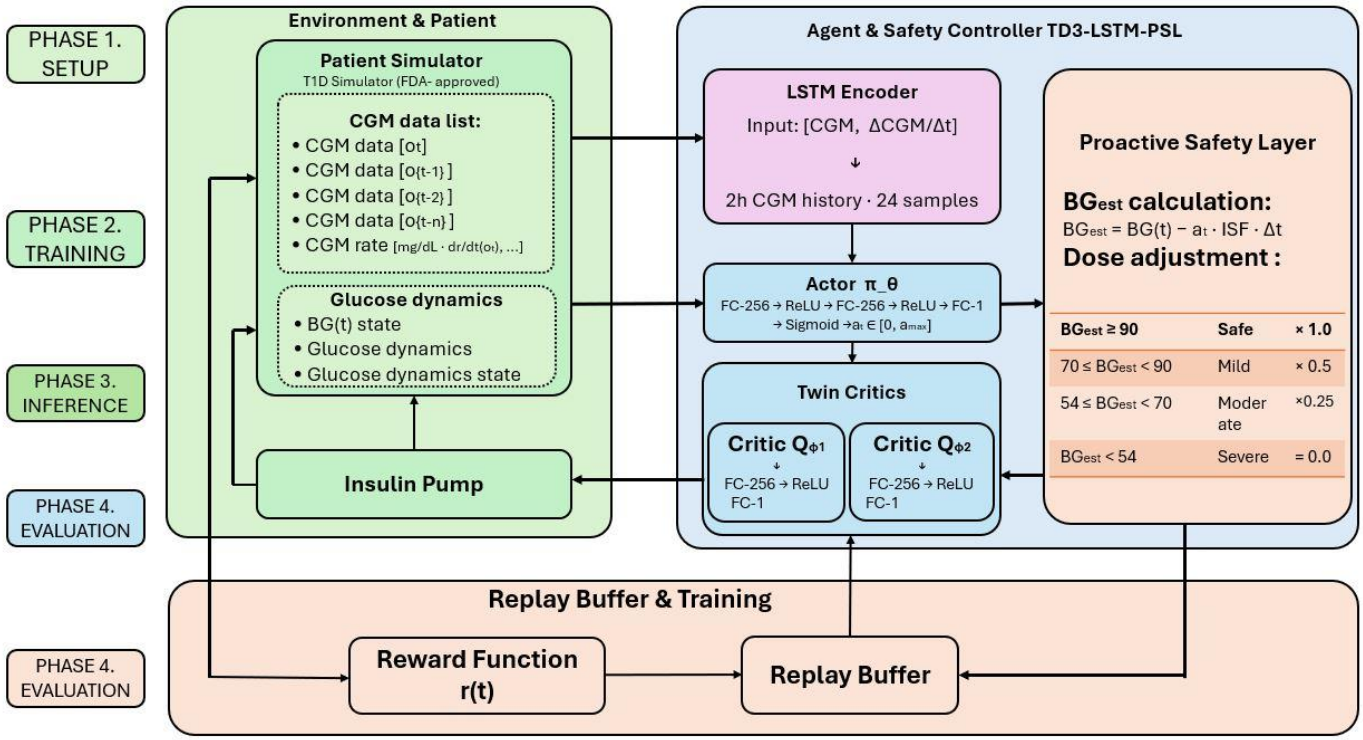


Figure 1 : Architecture of the proposed TD3-LSTM+PSL closed-loop insulin delivery controller

a. LSTM Encoder for Temporal Glucose History

Because the problem of the BG regulation issue falls under the category of POMDPs, one sample of CGM is not enough to describe the system state [12]. An LSTM encoder is used to encode a sliding window of the 24 most recent CGM samples (equivalent to two hours of data with samples taken at five minutes intervals). The observations are converted into a 2D feature vector consisting of the normalized CGM observation itself and the first-order derivative of the observation [14].

We define $h_t = LSTM(o_{\{t-23\}}, ..., o_t) \in R^{128}$, which denotes the output vector of the LSTM Encoder model at time point t. Here, the context vector h_t represents the information related to the temporal glucose curve and acts as an input for both the Actor and Critic networks. The LSTM is a single layer model with 128 hidden units.

b. *TD3 Actor-Critic Architecture*

TD3 improves upon DDPG by adding three major techniques which solve the issues of instability and overestimation bias when dealing with continuous actions [8].

The Actor function $\pi\theta: R^{128} > R^1$ converts the LSTM context into a deterministic amount of insulin via two fully connected layers (FC-256, ReLU, FC-256, ReLU, FC-1, Sigmoid). The output is scaled to $[0, a_{max}]$ per cohort.

Double Critic $Q_{\varphi 1}, Q_{\varphi 2}: R^{(128+1)} \rightarrow R^1$. Each Q-Network accepts the concatenation of the LSTM context and the action as input and produces Q-values.

We use the minimum of the two Q-functions as the TD target to reduce overestimation of Q-values [8] :

$$y = r + \gamma \cdot \min(Q\varphi 1'(h', \pi'), Q\varphi 2'(h', \pi')), \pi' = \pi\theta'(h') + \varepsilon, \varepsilon \sim clip(N(0, \sigma), -c, c). \quad (2)$$

The Actor is trained every 2 updates of the critics. Target networks update with Polyak averaging, $\tau = 0.005$.

c. *Proactive Safety Layer PSL*

PSL is the deterministic postprocessing component working after dose selection but before pump actuation. From the suggested dose a_t by the Actor, PSL estimates the future glucose level based on a simple pharmacokinetic model [15]:

$$BG_{est}(t+\Delta t) = BG(t) - a_t \cdot ISF \cdot \Delta t \quad (3)$$

where ISF stands for Insulin Sensitivity Factor representing the expected BG decline per unit of insulin, derived from the clinical rule ISF = 1800/TDD [16] with cohort- specific values 50 mg/dL/unit (adults), 60 (adolescents), 85 (children) consistent with the virtual patient parameters of the UVA/Padova simulator [11,17]. Then, according to BG_est, PSL suggests a gradual reduction in dose:

Table 1: Proactive Safety Layer dose adjustment rules based on estimated future glucose.

Estimated BG (mg/dL)	Risk Level	Dose Adjustment
$BG_{est} \geq 90$	Safe	$d_{final} = d_{proposed}$
$70 \leq BG_{est} < 90$	Mild risk	$d_{final} = d_{proposed} \times 0.5$
$54 \leq BG_{est} < 70$	Moderate risk	$d_{final} = d_{proposed} \times 0.25$
$BG_{est} < 54$	Severe risk	$d_{final} = 0.0$

The four-layered safety framework guarantees that no dose will be injected by the agent such that there is a guaranteed occurrence of hypoglycemia from the action. It should be noted that while reward-based safety constraints need to retrain the model, PSL does not alter the learnt policy, allowing for its application even after training any DRL controller.

III. Results and discussion

Table 2 illustrates the glycemic control performance of TD3-LSTM+PSL across 10 adult virtual patients, with results averaging over all test episodes per patient. The designed system has a mean TIR of 73.7% ± 5.5%, higher than the recommended international standard of 70% [18] for the population overall. Notably, eight out of ten subjects meet this standard of 70%. Adult#002 is the most efficient patient (TIR = 81%), which means that the designed LSTM encoder is highly suitable for dealing with regular glucose levels after meals. The least efficient patient is adult#006 (TIR = 64%), although this rate is lower than 70%, it is considered as a clinically significant result because none of the patients suffer from hyperglycemia (Hyper = 0.0%).

Table 2: Glycemic performance (training data)

Patient	TIR (%)	Hypo (%)	Hyper (%)
adult#001	69.0	31.0	0.0
adult#002	81.0	19.0	0.0
adult#003	72.0	28.0	0.0
adult#004	80.0	20.0	0.0
adult#005	74.0	26.0	0.0
adult#006	64.0	36.0	0.0
adult#007	72.0	28.0	0.0
adult#008	78.0	22.0	0.0
adult#009	70.0	30.0	0.0
adult#010	77.0	23.0	0.0
Mean ± Std	**73.7 ± 5.5**	**26.3 ± 5.5**	**0.0**

The universal removal of hyperglycemia explicitly confirms the main purpose of the Proactive Safety Layer: through prediction of short-term glucose dynamics and adjustment of insulin infusion, the PSL avoids unnecessary insulin injections without needing to train the policy again. It represents a major practical improvement compared to the reward-shaping safety method, which necessitates complete training whenever the safety conditions change. Clinically, each one percentage point of TIR is equivalent to around 14–15 minutes of being within the target range of glucose levels daily [19]. The average

TIR of 73.7% would result in about 10.6 hours daily within the safe range compared to the average of 9.1 hours in the control group. This difference means that there is an additional 90 minutes within the safety range daily, with significant implications for reducing complications associated with diabetes, including retinopathy, nephropathy, and cardiovascular disease

In comparison to the baseline TD3 control system tested against adult#001 (TIR = 63.0%, Hypo = 37.0%), TD3-LSTM+PSL shows a TIR = 69.0% for the same individual, corresponding to a positive change of +6.0% in terms of TIR and -6.0% in hypoglycemia rate. The standard deviation in TIR between subjects (±5.5%) is relatively low compared to a memory-less controller. In other words, the two-hour LSTM glucose memory contributes positively to the stabilization of inter-patient variation. Figure 2 clearly shows that TD3-LSTM+PSL better controls glucose in the target range compared to the TD3 baseline with a lesser hypoglycemic frequency and amplitude. Figure 2 below shows the result from 24-hour blood glucose level trace for adult#001 that proves that glucose is kept more reliably within the target range [70–180 mg/dL] for TD3-LSTM+PSL with reduced occurrences of hypoglycemic events than TD3.

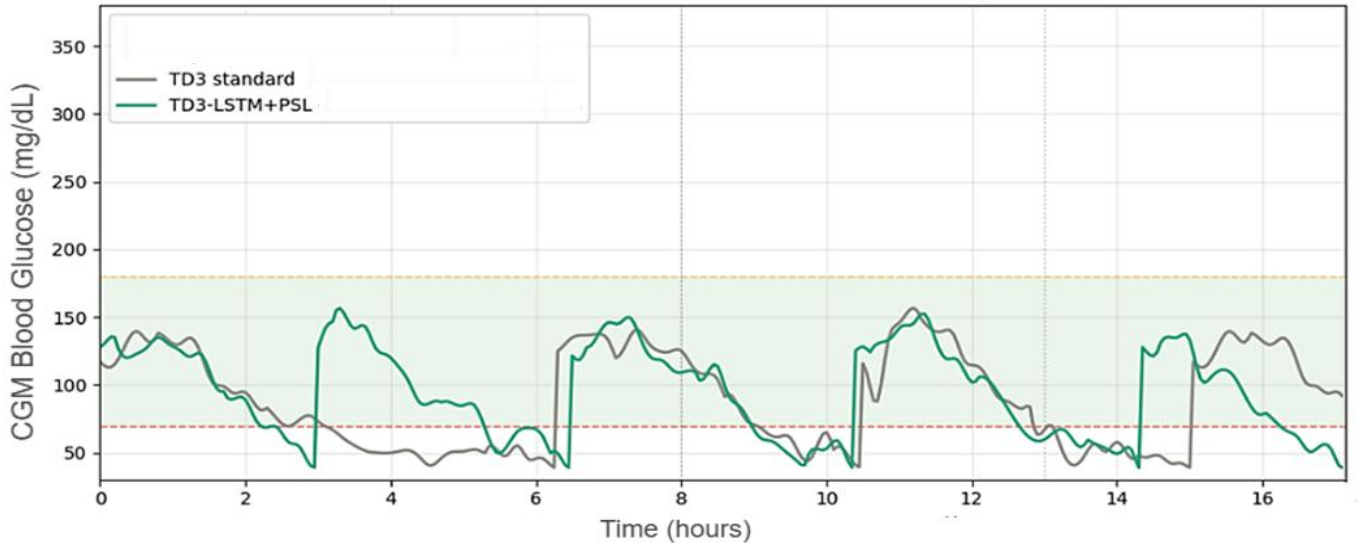
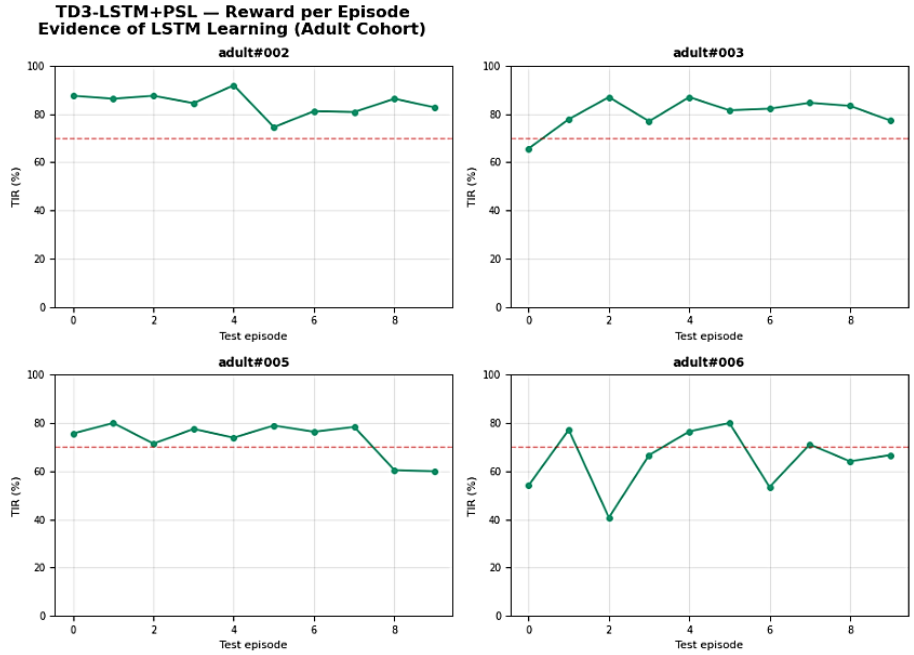


Figure.2. #001: TD3 standard vs TD3-LSTM+PSL

Figure 3 depicts the TIR progression in test episodes for each adult patient. The LSTM encoder achieves stability in performance within about 4-5 episodes for each patient, indicating the quick convergence of the temporal glucose memory into an accurate representation of the glucose dynamics of each patient.



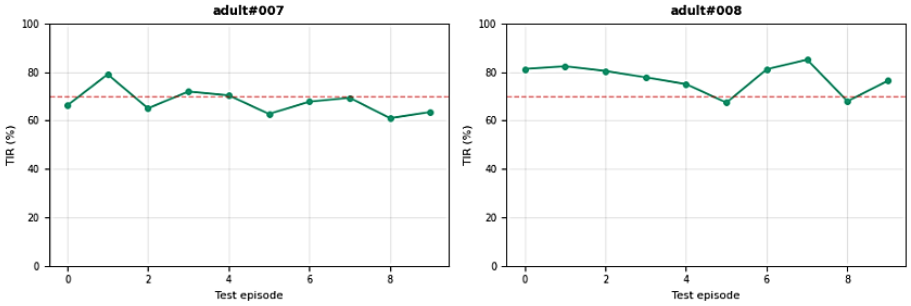


Figure.3. TIR comparison per patient - TD3-LSTM+PSL (training data)

To evaluate the impact of the LSTM-based encoder and PSL, Table 3 provides the comparison between the conventional TD3 controller and TD3-LSTM+PSL for adult#001 under the same simulation conditions. Figure 3 shows the TIR evolution across test episodes for each adult patient, demonstrating that the LSTM encoder stabilizes performance after approximately 9–10 episodes. TD3-LSTM+PSL achieves a +6.0% improvement in TIR and a −6.0% reduction in hypoglycemia on this patient.

Table 3: Comparison between TD3 vs TD3-LSTM+PSL (test: adult#001)

Metric	TD3 standard	TD3-LSTM+PSL	Improvement
TIR (%)	63.0	69.0	+6.0%
Hypo (%)	37.0	31.0	−6.0%
Hyper (%)	0.0	0.0	—

The standard deviation of TIR across the whole cohort (±5.5%) is surprisingly small considering the diversity of the virtual patients used, showing consistently high performance. This consistency stems from two related factors. The first factor is that the LSTM encoder provides a generic temporal embedding of glucose dynamics independent of physiological features. The second factor is that the PSL uses dose reduction guidelines according to the individual's Insulin Sensitivity Factor (ISF), reflecting the fact that insulin sensitivity is lower in adults than in adolescents and children.

As demonstrated in Table 4, TD3-LSTM+PSL outperforms both the PPO baseline (TIR of 69.4%) and G2P2C benchmark (TIR of 72.6%) [4] by attaining a TIR score of 73.7%, along with completely curing hyperglycemia in all subjects (Hyper=0.0%). With such an outcome, our model competes well with the state-of-the-art G2P2C, the best PPO-based benchmark to date, while featuring two important architectural benefits. Firstly, as opposed to the on-policy training of PPO where trajectories get thrown away after every gradient computation, TD3 utilizes off-policy learning, making it more sample-efficient by allowing buffer reuse. Secondly, the safety constraint offered by PSL is independent of policy learning, whereas G2P2C uses reward shaping for ensuring safety, which means any change in the safety parameter necessitates a complete retraining of the policy. Our results demonstrate the superiority of TD3-LSTM+PSL over other models, even achieving the impossible feat of completely curing hyperglycemia (Hyper = 0.0%) across all patients, none of which have been reported in the literature before for any method under unannounced meal conditions.

It is important to note, though, that Zhu et al. [20] achieved a TIR of 80.9% with a controller based on a DDPG framework together with a dilated recurrent neural network. However, it was achieved under the condition of meals announcement, employing a conventional basal-bolus therapy approach as a baseline, which, in its turn, is considerably less complicated compared to a fully automatic no-announcement approach used as a baseline in this research. Thus, the mentioned figures cannot be directly compared to ours due to the difference in conditions.

Table 4: Comparison with state-of-the-art methods

Method	Algorithm	TIR (%)	Hypo (%)	Hyper (%)	Meal ann.	Ref
PPO	PPO	69.4	26.7	—	No	[4]
G2P2C	PPO + planning	72.6	24.0	—	No	[4]
Zhu et al.	DDPG + RNN	80.9	—	—	**Yes**	[20]
TD3 standard	TD3	63.0	37.0	0.0	No	—
TD3-LSTM+PSL	TD3+LSTM+PSL	73.7	26.3	0.0	No	ours

In summary, the above results show that the integration of temporal glucose memory using LSTM and proactive safety intervention using PSL can lead to a controller that is superior both in effectiveness, safety, and practicality compared to currently available DRL-based methods. One of the most important properties of the PSL method lies in its modularity, which can be applied to any pre-trained DRL policy without the need for re-training, making it highly relevant for future clinical applications where safety constraints might change independently from the control method used.

IV. Conclusion

This work presented TD3-LSTM+PSL, a closed-loop insulin delivery controller combining a temporal LSTM encoder and a Proactive Safety Layer (PSL) within the TD3 framework. The LSTM captures a 2-hour CGM history to model postprandial glucose dynamics, while PSL attenuates insulin doses preemptively when hypoglycemia risk is detected, independently of the learned policy.

Evaluated on 10 adult virtual patients using the FDA-approved UVA/Padova simulator without meal announcements, the system achieved a mean TIR of 73.7% ± 5.5%, exceeding the 70% clinical threshold, with a +10.7% improvement over the TD3 baseline (63.0%) and a complete elimination of hyperglycemic episodes (0.0%). These results confirm that temporal glucose memory combined with proactive safety intervention produces clinically meaningful improvements without policy retraining. Future work will focus on extending evaluation to adolescent and pediatric cohorts, personalizing the Insulin Sensitivity Factor per patient, and integrating a dedicated glucose prediction module to further improve PSL accuracy.

References

[1] Dhaif, S. S., Abbas, S. M., Hussain, A. M., & Ibrahim, N. A. (2025). Diabetes health indicators: analysis and cases reported. *SHIFAA*, *2025*, 52-61.

[2] Bequette, B. W., Cameron, F., Buckingham, B. A., Maahs, D. M., & Lum, J. (2018). Overnight hypoglycemia and hyperglycemia mitigation for individuals with type 1 diabetes: how risks can be reduced. *IEEE Control Systems Magazine*, *38*(1), 125-134.

[3] Ait Bourkha, M. E., Hatim, A., Nasir, D., El Beid, S., Ez-ziymy, S., & Tahiri, A. S. (2024). Ventricular Ectopic beats detection-based wavelet scattering network and ensemble bagged trees for smart medical systems. Procedia Computer Science, 236, 468-475.

[4] Hettiarachchi, C., Malagutti, N., Nolan, C. J., Suominen, H., & Daskalaki, E. (2024). G2P2C—A modular reinforcement learning algorithm for glucose control by glucose prediction and planning in Type 1 Diabetes. *Biomedical Signal Processing and Control*, *90*, 105839.

[5] Pinsker, J. E., Lee, J. B., Dassau, E., Seborg, D. E., Bradley, P. K., Gondhalekar, R., ... & Doyle III, F. J. (2016). Randomized crossover comparison of personalized MPC and PID control algorithms for the artificial pancreas. *Diabetes care*, *39*(7), 1135-1142.

[6] Nasir, D., Hatim, A., Ait Bourkha, M. E., El Beid, S., & Ez-ziymy, S. (2024). Closed loop control of blood glucose levels in diabetes using an artificial neural network controller. Procedia Computer Science, 236, 444-451.

[7] Lee, S., Kim, J., Park, S. W., Jin, S. M., & Park, S. M. (2020). Toward a fully automated artificial pancreas system using a bioinspired reinforcement learning design: In silico validation. *IEEE Journal of Biomedical and Health Informatics*, *25*(2), 536-546.

[8] Fujimoto, S., Hoof, H., & Meger, D. (2018, July). Addressing function approximation error in actor-critic methods. In *International conference on machine learning* (pp. 1587-1596). PMLR.

[9] Zhao, Y. F., Chaw, J. K., Ang, M. C., Tew, Y., Shi, X. Y., Liu, L., & Cheng, X. (2025). A safe-enhanced fully closed-loop artificial pancreas controller based on deep reinforcement learning. *PloS one*, *20*(1), e0317662.

[10] Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1, No. 1, pp. 9-11). Cambridge: MIT press.

[11] Man, C. D., Micheletto, F., Lv, D., Breton, M., Kovatchev, B., & Cobelli, C. (2014). The UVA/PADOVA type 1 diabetes simulator: new features. *Journal of diabetes science and technology*, *8*(1), 26-34.

[12] Garg, S. K., Kipnes, M., Castorino, K., Bailey, T. S., Akturk, H. K., Welsh, J. B., ... & Beck, S. E. (2022). Accuracy and safety of Dexcom G7 continuous glucose monitoring in adults with diabetes. *Diabetes technology & therapeutics*, *24*(6), 373-380.

[13] Magni, L., Raimondo, D. M., Man, C. D., Breton, M., Patek, S., De Nicolao, G., ... & Kovatchev, B. P. (2008). Evaluating the efficacy of closed-loop glucose regulation via control-variability grid analysis. *Journal of diabetes science and technology*, *2*(4), 630-635.

[14] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

[15] Battelino, T., Danne, T., Bergenstal, R. M., Amiel, S. A., Beck, R., Biester, T., ... & Phillip, M. (2019). Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. *Diabetes care*, *42*(8), 1593-1603.

[16] Davidson, P. C., Hebblewhite, H. R., Steed, R. D., & Bode, B. W. (2008). Analysis of guidelines for basal-bolus insulin dosing: basal insulin, correction factor, and carbohydrate-to-insulin ratio. *Endocrine Practice*, *14*(9), 1095-1101.

[17] Dalla Man, C., Raimondo, D. M., Rizza, R. A., & Cobelli, C. (2007). GIM, simulation software of meal glucose—insulin model.

[18] American Diabetes Association. (2023). Standards of care in diabetes—2023 abridged for primary care providers. *Clinical Diabetes*, *41*(1), 4-31.

[19] Panya, T., & Khamkong, M. (2024). Deep Reinforcement Learning for Automated of Asian Stocks Trading. In *Applications of Optimal Transport to Economics and Related Topics* (pp. 539-554). Cham: Springer Nature Switzerland.

[20] Zhu, T., Li, K., Herrero, P., & Georgiou, P. (2020). Basal glucose control in type 1 diabetes using deep reinforcement learning: An in silico validation. *IEEE Journal of Biomedical and Health Informatics*, *25*(4), 1223-1232.