Article

Reinforcement Learning for Multiple Daily Injection (MDI) Therapy in Type 1 Diabetes (T1D)

Mehrad Jaloli [ORCID] and Marzia Cescon *

Department of Mechanical Engineering, University of Houston, Houston, TX 77004, USA; mjaloli@uh.edu
* Correspondence: mcescon2@central.uh.edu; Tel.: +1-713-743-2037

Abstract: In this study, we propose a closed-loop insulin administration framework for multiple daily injection (MDI) treatment using a reinforcement learning (RL) agent for insulin bolus therapy. The RL agent, based on the soft actor–critic (SAC) algorithm, dynamically adjusts insulin dosages based on real-time glucose readings, meal intakes, and previous actions. We evaluated the proposed strategy on ten in silico patients with type 1 diabetes undergoing MDI therapy, considering three meal scenarios. The results show that, compared to an open-loop conventional therapy, our proposed closed-loop control strategy significantly reduces glucose variability and increases the percentage of time the glucose levels remained within the target range. In particular, the weekly mean glucose level reduced from 145.34 ± 57.26 mg/dL to 115.18 ± 7.93 mg/dL, 143.62 ± 55.72 mg/dL to 115.28 ± 8.11 mg/dL, and 171.63 ± 49.30 mg/dL to 143.94 ± 23.81 mg/dL for Scenarios A, B and C, respectively. Furthermore, the percent time in range (70–180 mg/dL) significantly improved from $63.77 \pm 27.90\%$ to $91.72 \pm 9.27\%$ ($p = 0.01$) in Scenario A, $64.82 \pm 28.06\%$ to $92.29 \pm 9.15\%$ ($p = 0.01$) in Scenario B, and $58.45 \pm 27.53\%$ to $81.45 \pm 26.40\%$ ($p = 0.05$) in Scenario C. The model also demonstrated robustness against meal disturbances and insulin sensitivity disturbances, achieving mean glucose levels within the target range and maintaining a low risk of hypoglycemia, which were statistically significant for Scenarios B and C. The proposed model outperformed open-loop conventional therapy in all scenarios, highlighting the potential of RL-based closed-loop insulin administration models in improving diabetes management.

Keywords: blood glucose management; reinforcement learning; multiple daily injection therapy; simulation for clinical application; in silico validation

1. Introduction

Diabetes mellitus (DM) is a chronic metabolic disease defined by glucoregulatory dysfunction, resulting in hyperglycemia caused by either absolute insulin deficiency (type 1 diabetes or T1D) or relative insulin insufficiency and reduced sensitivity (type 2 diabetes or T2D) [1]. To prevent hyperglycemia and its related consequences, individuals diagnosed with T1D are dependent on exogenous insulin administration to control their blood glucose (BG) levels, while T2D individuals may need insulin treatment only at some different points during the progression of the disease [2]. However, the majority of T1D patients fail to achieve their glycemic goals, which makes it necessary to investigate proper insulin injection protocols in order to enhance the health and quality of life of these patients. Furthermore, the inter-patient variability of this disease necessitates personalized insulin administration techniques for the efficient management of glycemic control in people with DM [3].

One common strategy is the basal–bolus method, which is widely used to maintain BG levels within the euglycemic range of 70–180 mg/dL. In this approach, basal insulin regulates fasting BG levels, i.e., during night or intervals between two meals, while bolus doses compensate for the increase in BG concentration induced by meal intakes [4]. The

basal–bolus method can be implemented through either multiple daily injection (MDIs) [5] or continuous subcutaneous insulin infusions (CSII) using a pump [6].

It is noteworthy that most insulin-treated patients still prefer using MDI over insulin pumps, as the latter bring burdens associated with the use of an additional medical device attached to the body, insulin infusion set failures, and difficult access for some patients [7].

Accordingly, in this study, we focus on the development of a closed-loop insulin delivery system implementing MDI therapy for people with T1D. From the perspective of control theory, MDI therapy can be considered as an optimization problem in which basal and bolus insulin doses are chosen based on glucose, which are typically acquired by CGM sensors at a sampling rate of 5 to 15 min, in order to maintain BG levels within the target range. In general, basal doses are determined heuristically based on the patient's medical data and clinical experience by the physician, while meal boluses are calculated as follows:

$$u_{bolus} = \frac{CHO}{CR} + \frac{G_c - G_d}{CF} - u_{IOB}, \tag{1}$$

where u_{bolus} denotes the amount of bolus insulin, CHO stands for the estimated carbohydrate content of the meal, G_c and G_d represent the actual and target blood glucose level, respectively, u_{IOB} is the remaining insulin on board (IOB), and CR and CF are personalized insulin-to-carbohydrate ratio and correction factor, respectively.

There have been extensive studies in the diabetes management literature on developing insulin delivery algorithms to automate the MDI regimen using methodologies such as PID and fuzzy logic [8,9], optimization-based approaches [10–12], and iterative learning strategies [13–16].

With the rise in popularity of continuous glucose monitoring (CGM) sensors, a large amount of personal health data from the daily use of diabetes-related devices has been accumulated, which has led to a transformation in diabetes management and the development of data-driven and artificial intelligence (AI)-based approaches in more recent studies [17–19].

Machine learning (ML) and neural networks (NN) are subfields of AI that are widely used to detect and learn patterns from datasets and make accurate predictions and decisions [20–22]. In the field of diabetes management, several studies have employed these methods for insulin dosage recommendations [23,24] and future blood glucose (BG) predictions [25]. These models consider various patient-related physiological information, such as heart rate (HR) [26] and electrodermal activity (EDA) [27], in addition to insulin doses and meal intakes, to account for the factors that may affect BG variation, such as stress and physical activity [28,29].

However, data-driven techniques are highly dependent on the quality of recorded data. Since many datasets are collected under daily living conditions, they often contain missing points and errors, which can negatively impact the model's performance. To create a generalized data-driven model that provides optimal insulin doses to maintain normoglycemia under various conditions and for different patients, experiments need to be designed to examine BG variation in response to a wide range of disturbances, such as different insulin doses and carbohydrate intake at different times of the day, to simulate free-living conditions. Such experiments may pose a significant risk to patients' health.

To overcome these limitations, several studies have used in silico simulators to manipulate the factors that affect BG dynamics and create a generalized insulin delivery model. For instance, Ref. [30] proposed a k-nearest-neighbors (KNN)-based decision support system to provide weekly insulin dose recommendations, including basal, correction factor (CF), or carbohydrate ratio (CR) modification, using data obtained from a simulator established in [31]. Additionally, Cappon et al. [32] utilized the data generated by UVa/Padova, an FDA-approved BG simulator to train a NN that integrates with a standard bolus calculator and provides optimal correction doses for different meal sizes and pre-prandial glucose levels.

In recent years, reinforcement learning (RL) has gained a lot of attention among ML approaches. RL is a self-learning ML framework that enables an agent to learn how to make decisions in a closed-loop fashion, by interacting with an environment, receiving rewards or punishments based on its actions, and adjusting its behavior accordingly [33]. Through this process of trial and error, the agent can learn to make optimal decisions in complex and uncertain environments. RL has been successfully used in many applications such as gaming [34] and robotics [35]. More comprehensive descriptions regarding RL models and structures may be found in [33].

In recent years, RL has also been used in healthcare problems where optimal decisions are very crucial, such as treatment recommendations [36] or automated insulin dosage recommendation systems. In 2020, Zhu et al. [37] proposed a deep RL model for basal insulin and glucagon delivery. Since, RL can learn the task of an optimal control, it can be used for the insulin decision in MDI by continuously optimizing the rewards in terms of glucose outcomes to determine the insulin dosage within a long control period (e.g., one day or one week) [38].

However, one of the drawbacks of using RL is that it cannot handle constraints well and generally requires a long trial-and-error process for agent training. In other words, the actions taken by the agent, specifically in the initial steps of the training process, would cause serious complications for the environment, which in our case is the patient's body. This is a critical factor that limits the application of RL in real life and clinical cases. In recent studies, various RL approaches, such as actor–critic (AC), Q-learning, SARSA, and Gaussian process reinforcement learning, have been examined for glucose control solely utilizing in silico datasets.

Accordingly, to deal with this issue, prior to its implementation in clinical studies on humans, in silico simulations assist the development of RL systems and their verification under several virtual experimental conditions.

In this study, we propose an insulin delivery system for T1D patients that incorporates a Soft Actor–Critic (SAC) RL model for meal bolus insulin dose suggestion and an iterative learning controller (ILC) for long-acting insulin administration, as suggested in [5]. To evaluate the feasibility and effectiveness of our proposed model for the simulated MDI therapy, we used a mathematical model of the glucose–insulin metabolic system that incorporates a pharmacokinetics (PK) module for long-acting basal insulin proposed in [5]. In the following study, we show that the integration of the SAC RL model and the ILC controller has the potential to improve glycemic control and reduce the burden of diabetes management for T1D patients.

2. Materials and Methods

In this section, an MDI treatment framework is constructed by integrating an RL agent into the simulator model and training the model using the real-time data of 10 in- silico T1D adults generated by the simulator. It should be noted that the patients' individual parameters were extracted from [39,40]. The pre-trained RL agent is then used in conjunction with a controller to construct an automated closed-loop insulin administration model with MDI treatment. The block diagram of the proposed framework is shown in Figure 1.
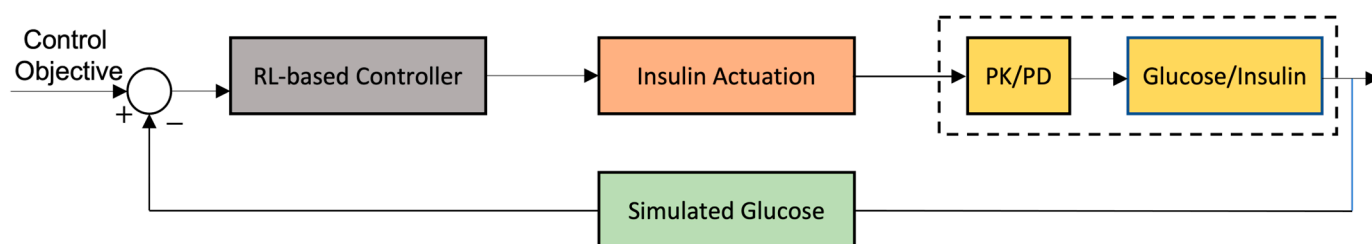

Figure 1. The block diagram of the proposed closed-loop RL-based controller for MDI therapy for virtual T1D patients.

2.1. Simulator

A metabolic model of a virtual patient able to simulate MDI treatment in people with T1D was first proposed by Cescon et al. in [5]. By integrating the SAC RL [41] model, we aim to create a closed-loop insulin delivery system that can dynamically adjust insulin dosages based on real-time glucose readings. Figure 2 depicts the SAC RL agents integrated into a metabolic model.
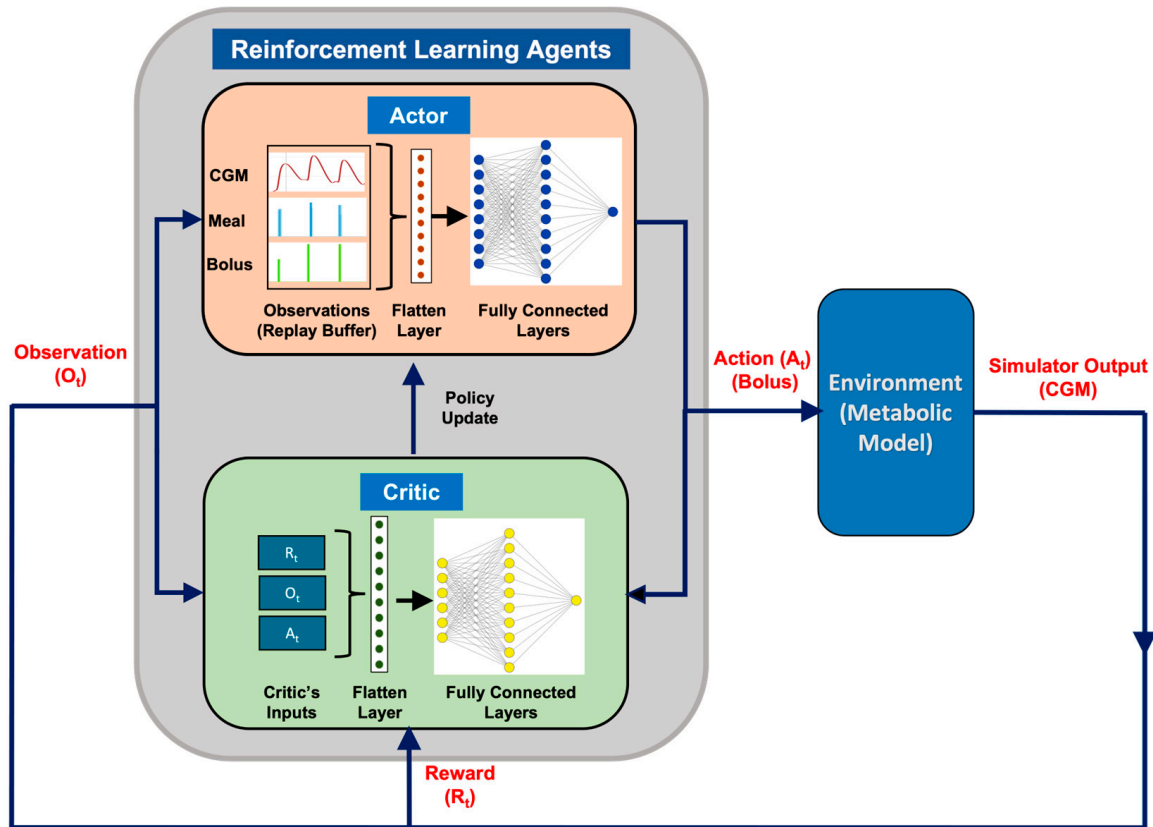

Figure 2. The proposed framework for MDI therapy using an SAC RL methodology. Reinforcement learning (RL) agent is integrated into a metabolic model to dynamically adjust insulin dosages based on real-time glucose readings generated by the simulator.

The SAC RL model is a variant of the actor–critic algorithm that is designed to learn a stochastic policy that maximizes the expected cumulative reward. The policy is represented as a function of the current state, and it is learned through an actor–critic approach where the actor learns the policy and the critic learns the state-value function [41]. The SAC RL model uses a maximum entropy formulation to optimize the policy, which encourages exploration and prevents premature convergence to suboptimal policies. The objective function for SAC involves maximizing the expected reward and the entropy of the policy, and the critic network is trained to minimize the mean squared error between the predicted state-value function and the actual state-value function [41]. The integration of the SAC RL model into the metabolic model [5] involves training the RL agent using real-time simulated data and then using the trained model to automate insulin delivery inside the simulator. The RL agent learns to provide bolus insulin dosages based on real-time glucose readings, meal intakes and previous action taken by the agents, according to an optimal policy π that maximizes an objective function $J(\pi)$:

$$J(\pi) = E_{\{\tau \sim \pi\}} \left[\sum_{\{t=0\}}^{T} \gamma^t (r(s_t, a_t) + \alpha H(\pi(a_t|s_t))) \right], \qquad (2)$$

where τ is a trajectory sampled from π, s_t is the state at time t, a_t is the action taken at time t, $r(s_t, a_t)$ is the reward function, and $H(\pi(a_t|s_t))$ is the entropy of the policy at state s_t and action a_t. T is the time horizon, γ is the discount factor, and α is a hyperparameter that controls the trade-off between the expected reward and the entropy of the policy.

2.2. Reward Function

The bolus advisor that we propose aims to maximize the time within the range of 70–180 mg/dL which leads to reducing the percentage of time of the day patients experience hypoglycemia (BG < 70 mg/dL) or hyperglycemia (BG > 180 mg/dL). For this, the reward function takes the previous BG value and meal information generated by the simulator, as well as an array containing actions taken in the last three hours by the agent, and two constant values representing the desired BG range, which are 70 and 180 mg/dL. BG readings were obtained at 1 min sample rate. However, to achieve a model that is capable of providing insulin dosage for MDI therapy, the agent needs to take actions only at mealtime. Accordingly, we set the sampling rate of the RL agent to 15 min and defined the reward function such that if the agent takes action at the immediate time stamps after mealtimes, a positive reward is introduced; otherwise, a negative reward as a punishment is obtained if the agent takes action in any other time. The rewards are generated every 15 min, equal to the sampling rate of the RL agent, and then are summed up at the end of each episode, which lasts one week. Algorithm 1 presents the pseudocode for the reward function, providing further details of its implementation.

Algorithm 1 Reward Function

Initialize parameters BG_t, BG^{buffer}, $Meal^{buffer}$, $Action^{buffer}$, BG^{bounds}
Inputs:
    BG_t: Current continuous glucose monitor (CGM) readings
    BG^{buffer}: A buffer of CGM readings for the past 3 h.
    $Meal^{buffer}$: A buffer of previous meal intakes for the past 3 h.
    $Action^{buffer}$: A buffer of previous insulin doses for the past 3 h.
    BG^{bounds}: Bounds of the BG readings, i.e., (70, 180) mg/dL
Output:
    $r^{episode}$: Calculated reward for each episode
$BG^{Target} \leftarrow 125$
$BG_{t-1} = BG_1^{buffer}$
$Meal_{t-1} = Meal_1^{buffer}$
$Action_{t-1} = Action_1^{buffer}$
for each episode do
    for each step, $t = 1, \ldots, N, N$ being the number of steps in each episode do
        if $Action_{t-1} > 0$ and $Meal_{t-1} > 0$ then
            $r_t^{Action}: \leftarrow 10$
        else if $Action_{t-1} \leq 0$ and $Meal_{t-1} \leq 0$ then
            $r_t^{Action}: \leftarrow 0$
        else
            $r_t^{Action} :\leftarrow -2$
        end if
        if $BG_t \geq \mathbf{min}BG^{bounds}$ and $BG_t \leq \mathbf{max}BG^{bounds}$ then
            $r_t^{BG}: \leftarrow 0.1 \times \mathbf{exp}(-|BG_t - BG^{Target}|)/100$
        else
            $r_t^{BG} \leftarrow -0.01 \times (BG_t - BG^{Target}|)$
        end if
        r (t) $\leftarrow r_t^{BG} + r_t^{Action}$
    end for
    $r^{episode} = \sum_1^N r(t)$
end for

2.3. Simulation Scenarios

To evaluate the performance of the proposed model under different conditions, three meal scenarios were considered:

- **Scenario A (nominal case).** Each day, meals were at 7 a.m., 1 p.m., and 7 p.m., and the quantity of carbohydrate consumed each meal was 50 g, 75 g, and 75 g, for breakfast, lunch, and dinner, respectively. The rationale behind using this scenario was to evaluate the controller's ability to maintain stable blood glucose levels under nominal conditions.

- **Scenario B (robustness against meal disturbances).** Meal timings for each meal were normally distributed, with means 30 min and standard deviation of 3 min. Further, the amount of carbohydrate for each meal was normally distributed with (50, 75, 75) g and standard deviations (5, 7.5, 7.5), for breakfast, lunch, and dinner, respectively. We used this scenario to test the controller's robustness against meal disturbances, such as variations in meal timing and size, which are common in real-world situations. These variations can have a significant impact on blood glucose levels, and the controller should be able to adapt to them to maintain stable levels.

- **Scenario C (robustness against insulin sensitivity disturbances).** Scenario B dictated the meal times and amounts of carbohydrates consumed. Moreover, to assess the robustness of the suggested algorithm against fluctuations in insulin sensitivity, insulin resistance was induced by modifying the parameters in the metabolic model that affect insulin's impact on glucose uptake and endogenous glucose production by 40%. The logic behind defining this scenario was to test the controller's robustness against changes in insulin sensitivity, which can occur due to various factors such as stress, illness, or changes in physical activity. Insulin resistance can make it challenging to maintain stable blood glucose levels, and the controller should be able to adjust the insulin dosage appropriately.

Overall, the three scenarios were chosen to evaluate the controller's performance under different conditions and to assess its robustness against various challenges that a real-world patient with type 1 diabetes might face.

3. Results

In this study, we utilized simulation to assess the effectiveness of the proposed methodology on ten in silico patients with type 1 diabetes undergoing MDI therapy for the three different scenarios previously described in Section 2. As far as long-acting insulin is concerned, we used the ILC controller proposed in [5]. On the other hand, for the short-acting insulin, the SAC RL agent was initially trained on simulated data from scenario B. Table 1 presents the specifications and hyperparameters of the developed model, which were optimized via grid search to obtain the optimal parameter values.

Table 1. Training hyperparameter and model specifications for the SAC RL model.

Hyperparameter/Specification	Value
Critic optimizer	Adam
Critic learning rate	1×10^{-4}
Critic gradient threshold	1
Critic L2 regularization factor	2×10^{-4}
Actor optimizer	Adam
Actor learning rate	1×10^{-4}
Actor gradient threshold	1
Actor L2 regularization factor	1×10^{-4}
Target smooth factor	1×10^{-3}
Mini batch size	512
Experience buffer length	1×10^{7}
Number of hidden units	256
Sample time	5 min

The pre-trained SAC RL agent was subsequently integrated into the simulator to enable dynamic bolus insulin dosage administration. The simulation protocol spanned 14 days, starting at 7:00 on Day 0. The performance of the reinforcement learning (RL) agent was evaluated solely on the second week, starting from Day 8.

The simulations started with initial conditions of 125 mg/dL, with no insulin present in the plasma or subcutaneous depot, and a reference basal glucose concentration of 125 mg/dL. The RL agents' performances were compared with open-loop therapy consisting of one long-acting insulin injection per day, set at 0.25 U per kg of body weight.

Table 2 presents the numerical results of the study, where weekly minimum, maximum, and mean glucose levels mg/dL, the percent time in range based on a blind continuous glucose monitoring (CGM) system, and the average daily rapid-acting insulin dosage U for each patient were computed. Population statistics such as mean and standard deviation for the mentioned metrics were computed separately for the RL and open-loop therapy and were arranged in different columns and rows based on the scenarios. The performance of the RL agent was compared to the open-loop conventional therapy using paired *t*-tests with a significance level of 5%. The open-loop model was used as a reference for evaluating the performance of the RL agent.

Table 2. Glycemic outcomes for the in silico population: mean $\pm$ standard deviation. Min, max, mean of generated CGM samples; BG % time in zones of blind CGM and average total daily rapid-acting insulin units.

	Metrics	Open Loop	RL	*p*-Value
Scenario A	min mg/dL	84.46 $\pm$ 40.22	69.06 $\pm$ 15.67	0.13
	max mg/dL	273.07 $\pm$ 131.21	179.20 $\pm$ 18.62	**0.03**
	mean mg/dL	145.34 $\pm$ 57.26	115.18 $\pm$ 7.93	0.07
	% time $\in$ (70, 180) mg/dL	63.77 $\pm$ 27.90	91.72 $\pm$ 9.27	**0.01**
	% time < 70 mg/dL	11.36 $\pm$ 24.98	5.68 $\pm$ 9.32	0.26
	% time < 54 mg/dL	7.70 $\pm$ 19.31	1.10 $\pm$ 2.40	0.16
	% time > 180 mg/dL	24.86 $\pm$ 27.65	2.58 $\pm$ 3.29	**0.02**
	% time > 250 mg/dL	5.22 $\pm$ 14.51	0	0.14
	Bolus U	17.92 $\pm$ 5.62	30.71 $\pm$ 31.47	0.08
Scenario B	min mg/dL	83.93 $\pm$ 40.32	69.74 $\pm$ 15.89	0.15
	max mg/dL	267.74 $\pm$ 126.51	181.92 $\pm$ 18.99	**0.04**
	mean mg/dL	143.62 $\pm$ 55.72	115.28 $\pm$ 8.11	0.08
	% time $\in$ (70, 180) mg/dL	64.82 $\pm$ 28.06	92.29 $\pm$ 9.15	**0.01**
	% time < 70 mg/dL	11.78 $\pm$ 26.11	5.49 $\pm$ 9.10	0.25
	% time < 54 mg/dL	7.82 $\pm$ 20.02	0.88 $\pm$ 1.93	0.15
	% time > 180 mg/dL	23.39 $\pm$ 26.43	2.21 $\pm$ 2.63	**0.02**
	% time > 250 mg/dL	4.64 $\pm$ 12.60	0	0.14
	Bolus U	18.50 $\pm$ 5.81	30.66 $\pm$ 31.31	0.09
Scenario C	min mg/dL	105.66 $\pm$ 27.06	97.24 $\pm$ 27.83	0.3
	max mg/dL	278.04 $\pm$ 86.77	209.83 $\pm$ 26.26	**0.03**
	mean mg/dL	171.63 $\pm$ 49.30	143.94 $\pm$ 23.81	0.1
	% time $\in$ (70, 180) mg/dL	58.45 $\pm$ 27.53	81.45 $\pm$ 26.40	**0.05**
	% time < 70 mg/dL	3.53 $\pm$ 11.18	0.026 $\pm$ 0.08	0.17
	% time < 54 mg/dL	1.20 $\pm$ 3.76	0	0.17
	% time > 180 mg/dL	38.00 $\pm$ 29.80	18.51 $\pm$ 26.40	0.11
	% time > 250 mg/dL	8.60 $\pm$ 19.10	0.24 $\pm$ 0.77	0.1
	Bolus U	17.92 $\pm$ 5.62	32.08 $\pm$ 31.28	0.06

The main goal of the study was to reduce glucose variability and increase the percentage of time the glucose levels remained within the target range of 70–180 mg/dL for all scenarios using the RL agent. The use of the RL agent resulted in a reduction in glycemic variability across the population such that the population average blood glucose (BG) levels reduced from 145.34 $\pm$ 57.26 mg/dL to 115.18 $\pm$ 7.93 mg/dL, 143.62 $\pm$ 55.72 mg/dL to 115.28 $\pm$ 8.11 mg/dL, and 171.63 $\pm$ 49.30 mg/dL to 143.94 $\pm$ 23.81 mg/dL for Scenarios A

to C, respectively. Additionally, the study found a significant decrease in the maximum blood glucose (BG) levels across all subjects in all three scenarios. Specifically, in Scenario A, the maximum BG levels decreased from 273.07 ± 131.21 mg/dL to 179.20 ± 18.62 mg/dL ($p = 0.03$), in Scenario B from 267.74 ± 126.51 mg/dL to 181.92 ± 18.99 mg/dL ($p = 0.04$), and in Scenario C from 278.04 ± 86.77 mg/dL to 209.83 ± 26.26 mg/dL ($p = 0.03$).

Moreover, the average percentage of time of the day that subjects experienced hyperglycemia (BG > 180 mg/dL) decreased in all scenarios, with reductions from $24.86 \pm 27.65\%$ to $2.58 \pm 3.29\%$ in Scenario A, $23.39 \pm 26.43\%$ to $2.21 \pm 2.63\%$ in Scenario B, and $38.00 \pm 29.80\%$ to $18.51 \pm 26.40\%$ in Scenario C. Notably, the reductions in Scenarios A and B were statistically significant ($p = 0.02$ for both scenarios).

In addition, the study revealed that the administration of short-acting insulin dosages by the RL agent resulted in a statistically significant improvement in the percentage of time glucose levels were within the target range for all scenarios. Specifically, the percentage of time within the target range increased from $63.77 \pm 27.90\%$ to $91.72 \pm 9.27\%$ ($p = 0.01$) in Scenario A, $64.82 \pm 28.06\%$ to $92.29 \pm 9.15\%$ ($p = 0.01$) in Scenario B, and $58.45 \pm 27.53\%$ to $81.45 \pm 26.40\%$ ($p = 0.05$) in Scenario C.

Figure 3 illustrates the performance of the proposed RL model compared to the open-loop simulator for Scenarios A to C, in terms of the populated median, 25th and 75th percentiles of the BG reading produced by the simulator in a 3-day scenario. It is observable that in all of the scenarios, the RL agent is capable of keeping the BG levels closer in the desired glycemic range (70, 180 mg/dL) compared to the open-loop case.
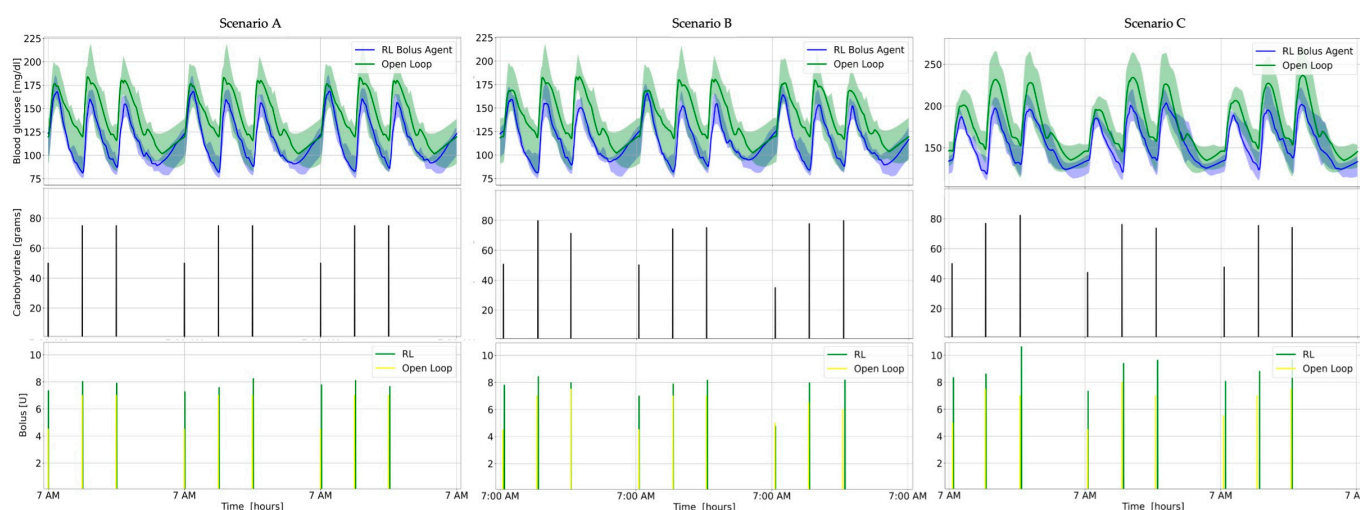

Figure 3. Performance of the proposed model compared to the open-loop simulator for Scenarios A to C, arranged from left to right, respectively. The top panel in each column presents the distribution of blood glucose (BG) values in terms of the median, as well as the 25th and 75th percentiles, of the BG population for 10 patients over 3 days, represented by blue and green colors for RL and open-loop models, respectively. The middle panel depicts the simulated meals, while the bottom panel shows the populated median bolus dose administered by the RL agent and open-loop model in yellow and green colors, respectively.

4. Discussion

In this study, a closed-loop insulin delivery system for MDI treatment of type 1 diabetes patients is proposed. The system uses a RL agent integrated into a metabolic model to dynamically adjust insulin dosages based on real-time glucose readings. The RL agent is trained using real-time simulated data and pre-trained for bolus insulin dosage administration. The performance of the RL agent is evaluated under three different scenarios with ten in silico patients with type 1 diabetes. The scenarios considered were nominal, robustness against meal disturbances, and robustness against insulin sensitivity disturbances.

The results obtained from the simulations show that the proposed RL agent outperformed the open-loop model in terms of reducing glucose variability and increasing the percentage of time the glucose levels remained within the target range. The proposed method demonstrated robustness against meal disturbances and fluctuations in insulin sensitivity, suggesting its potential for real-world applications. The results of the study demonstrate the potential of the proposed model to improve glucose control in patients with type 1 diabetes undergoing MDI therapy. The use of an automated closed-loop insulin administration model with MDI treatment, which integrates the SAC RL agent into the UVA/Padova simulator model, led to better glucose control in all three scenarios compared to open-loop therapy. Notably, the RL agent achieved statistically significant improvements in glucose control in scenarios B and C, which were designed to test the robustness of the proposed model against meal and insulin sensitivity disturbances. These findings suggest that the SAC RL model is capable of adapting to changing conditions and optimizing insulin dosages accordingly, making it a promising approach for improving glucose control in real-world settings.

It is worth noting that while simulation studies can provide valuable insights into the potential effectiveness of closed-loop insulin delivery systems, further research is needed to validate these findings in clinical settings. Future studies could involve testing the proposed model in clinical trials, where patients with type 1 diabetes could use the closed-loop system for an extended period of time and provide feedback on its usability and effectiveness. Additionally, it will be important to assess the safety and reliability of the system and identify any potential risks or limitations that may need to be addressed before widespread implementation. Overall, the proposed model represents a promising step towards improving glucose control and reducing the burden of diabetes management for patients undergoing MDI therapy.

5. Conclusions

In conclusion, we have presented a novel methodology that integrates a RL agent into a simulator model to automate insulin delivery in patients with type 1 diabetes undergoing MDI therapy. The proposed approach was evaluated using three different meal scenarios and was found to be effective in reducing glucose variability and increasing the percentage of time the glucose levels remained within the target range. The results demonstrated that the automated closed-loop insulin administration model with MDI treatment based on SAC RL agent is superior to open-loop therapy consisting of one long-acting insulin injection per day. This study provides a promising avenue for the development of automated insulin delivery systems that can dynamically adjust insulin dosages based on real-time glucose readings. Future research could focus on testing the proposed approach in clinical settings and exploring the use of other RL algorithms and control strategies to further improve the performance of the system. Ultimately, the goal is to provide patients with type 1 diabetes with a safe and effective way to manage their condition and improve their quality of life.

Incorporating the effect of other glycemic disturbances, such as physical activity (PA) and stress state (SS), to mention a few, is the focus of our current investigations. In order to do so, we are developing physiological models describing the effect of PA and SS to glucose metabolism, and once they are ready, we will proceed with the evaluation of a RL-based controller. We would like to stress that the main aim of this paper was to show the feasibility of utilizing RL-based methodologies in diabetes management. Future work will be devoted to the application of RL to more realistic scenarios.

The use of RL models in diabetes management has shown promising results in improving glycemic control in patients with T1DM. However, there are still some challenges that need to be addressed before these models can be widely adopted in clinical practice. One of the challenges is the need for a large amount of high-quality data to train the models. This requires collaboration between healthcare providers and data scientists to ensure that the data used to train the models is accurate, reliable, and representative of the patient population. Another challenge is the need for personalized RL models that can adapt to

the unique needs and preferences of each patient. This requires the development of more sophisticated algorithms that can learn from individual patient data and adjust treatment recommendations accordingly.

Despite these challenges, the application of RL models in diabetes management has the potential to significantly improve patient outcomes and reduce healthcare costs. By providing personalized treatment recommendations, these models can help patients achieve better glycemic control while minimizing the risk of hypoglycemia and other complications. They can also reduce the burden on healthcare providers by automating routine tasks and allowing them to focus on more complex aspects of diabetes care. As the technology continues to evolve and more data become available, RL models are likely to become an increasingly important tool in the management of diabetes and other chronic diseases.

References

1. American Diabetes Association Professional Practice (ADAPP) Committee. 2. Classification and diagnosis of diabetes: Standards of Medical Care in Diabetes—2022. *Diabetes Care* **2022**, *45*, S17–S38. [CrossRef] [PubMed]
2. American Diabetes Association (ADA). 9. Pharmacologic approaches to glycemic treatment: Standards of medical care in diabetes-2020. *Diabetes Care* **2020**, *43*, S98–S110. [CrossRef] [PubMed]
3. Solomon, T.P.J. Sources of inter-individual variability in the therapeutic response of blood glucose control to exercise in type 2 diabetes: Going beyond exercise dose. *Front. Physiol.* **2018**, *9*, 896. [CrossRef] [PubMed]
4. Boiroux, D.; Aradóttir, T.B.; Nørgaard, K.; Poulsen, N.K.; Madsen, H.; Jørgensen, J.B. An adaptive nonlinear basal-bolus calculator for patients with type 1 diabetes. *J. Diabetes Sci. Technol.* **2017**, *11*, 29–36. [CrossRef]
5. Cescon, M.; Deshpande, S.; Nimri, R.; Doyle, F.J., III.; Dassau, E. Using iterative learning for insulin dosage optimization in multiple-daily-injections therapy for people with type 1 diabetes. *IEEE Trans. Biomed. Eng.* **2020**, *68*, 482–491. [CrossRef]
6. Linkeschova, R.; Raoul, M.; Bott, U.; Berger, M.; Spraul, M. Less severe hypoglycaemia, better metabolic control, and improved quality of life in Type 1 diabetes mellitus with continuous subcutaneous insulin infusion (CSII) therapy; an observational study of 100 consecutive patients followed for a mean of 2 years. *Diabet. Med.* **2002**, *19*, 746–751. [CrossRef]
7. Tanenbaum, M.L.; Hanes, S.J.; Miller, K.M.; Naranjo, D.; Bensen, R.; Hood, K.K. Diabetes device use in adults with type 1 diabetes: Barriers to uptake and potential intervention targets. *Diabetes Care* **2017**, *40*, 181–187. [CrossRef]
8. Campos-Delgado, D.U.; Femat, R.; Hernández-Ordoñez, M.; Gordillo-Moscoso, A. Self-tuning insulin adjustment algorithm for type 1 diabetic patients based on multi-doses regime. *Appl. Bionics Biomech.* **2005**, *2*, 61–71. [CrossRef]
9. Campos-Delgado, D.U.; Hernández-Ordoñez, M.; Femat, R.; Gordillo-Moscoso, A. Fuzzy-based controller for glucose regulation in type-1 diabetic patients by subcutaneous route. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 2201–2210. [CrossRef]
10. Kirchsteiger, H.; Del Re, L.; Renard, E.; Mayrhofer, M. Robustness properties of optimal insulin bolus administrations for type 1 diabetes. In Proceedings of the 2009 American Control Conference, St. Louis, MO, USA, 10–12 June 2009; pp. 2284–2289.
11. Cescon, M.; Stemmann, M.; Johansson, R. Impulsive predictive control of T1DM glycemia: An in-silico study. In Proceedings of the Dynamic Systems and Control Conference, Fort Lauderdale, FL, USA, 17–19 October 2012; American Society of Mechanical Engineers: New York, NY, USA; Volume 45295, pp. 319–326.

12. Carrasco, D.S.; Matthews, A.D.; Goodwin, G.C.; Delgado, R.A.; Medioli, A.M. Design of MDIs for type 1 diabetes treatment via rolling horizon cardinality-constrained optimisation. *IFAC-PapersOnLine* **2017**, *50*, 15044–15049. [CrossRef]

13. Cescon, M.; Deshpande, S.; Doyle, F.J.; Dassau, E. Iterative learning control with sparse measurements for long-acting insulin injections in people with type 1 diabetes. In Proceedings of the 2019 American Control Conference (ACC), Philadelphia, PA, USA, 10–12 July 2019; pp. 4746–4751.

14. Owens, C.; Zisser, H.; Jovanovic, L.; Srinivasan, B.; Bonvin, D.; Doyle, F.J. Run-to-run control of blood glucose concentrations for people with type 1 diabetes mellitus. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 996–1005. [CrossRef]

15. Palerm, C.C.; Zisser, H.; Bevier, W.C.; Jovanovic, L.; Doyle, F.J., III. Prandial insulin dosing using run-to-run control: Application of clinical data and medical expertise to define a suitable performance metric. *Diabetes Care* **2007**, *30*, 1131–1136. [CrossRef]

16. Zisser, H.; Palerm, C.C.; Bevier, W.C.; Doyle, F.J., III.; Jovanovic, L. Clinical update on optimal prandial insulin dosing using a refined run-to-run control algorithm. *J. Diabetes Sci. Technol.* **2009**, *3*, 487–491. [CrossRef] [PubMed]

17. Woldaregay, A.Z.; Årsand, E.; Walderhaug, S.; Albers, D.; Mamykina, L.; Botsis, T.L.; Hartvigsen, G. Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. *Artif. Intell. Med.* **2019**, *98*, 109–134. [CrossRef] [PubMed]

18. Vettoretti, M.; Cappon, G.; Facchinetti, A.; Sparacino, G. Advanced diabetes management using artificial intelligence and continuous glucose monitoring sensors. *Sensors* **2020**, *20*, 3870. [CrossRef] [PubMed]

19. Contreras, I.; Vehi, J. Artificial intelligence for diabetes management and decision support: Literature review. *J. Med. Internet Res.* **2018**, *20*, e10775. [CrossRef]

20. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.

21. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

22. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [CrossRef]

23. Zhu, T.; Li, K.; Herrero, P.; Georgiou, P. Deep learning for diabetes: A systematic review. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 2744–2757. [CrossRef]

24. Bailey, T.S.; Walsh, J.; Stone, J.Y. Emerging technologies for diabetes care. *Diabetes Technol. Ther.* **2018**, *20*, S2–S78. [CrossRef]

25. Jaloli, M.; Cescon, M. Predicting Blood Glucose Levels Using CNN-LSTM Neural Networks. 2020 Diabetes Technology Meeting Abstracts. *J. Diabetes Sci. Technol.* **2021**, *15*, A432. [CrossRef]

26. Mirshekarian, S.; Shen, H.; Bunescu, R.; Marling, C. LSTMs and neural attention models for blood glucose prediction: Comparative experiments on real and synthetic data. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 706–712.

27. Marling, C.; Bunescu, R. The OhioT1DM dataset for blood glucose level prediction: Update 2020. In Proceedings of the CEUR Workshop Proc, Siena, Italy, 16–20 November 2020; Volume 2675, p. 71.

28. Jaloli, M.; Cescon, M. Demonstrating the Effect of Daily Physical Activities on Blood Glucose Level Variation in Type 1 Diabetes. *Diabetes Technol. Ther.* **2022**, *24*, A79, MARY ANN LIEBERT, INC 140 HUGUENOT STREET, 3RD FL, NEW ROCHELLE, NY 10801 USA.

29. Jaloli, M.; Lipscomb, W.; Cescon, M. Incorporating the Effect of Behavioral States in Multi-Step Ahead Deep Learning Based Multivariate Predictors for Blood Glucose Forecasting in Type 1 Diabetes. *BioMedInformatics* **2022**, *2*, 715–726. [CrossRef]

30. Tyler, N.S.; Mosquera-Lopez, C.M.; Wilson, L.M.; Dodier, R.H.; Branigan, D.L.; Gabo, V.B.; Guillot, F.H.; Hilts, W.W.; El Youssef, J.; Castle, J.R.; et al. An artificial intelligence decision support system for the management of type 1 diabetes. *Nat. Metab.* **2020**, *2*, 612–619. [CrossRef] [PubMed]

31. Resalat, N.; El Youssef, J.; Tyler, N.; Castle, J.; Jacobs, P.G. A statistical virtual patient population for the glucoregulatory system in type 1 diabetes with integrated exercise model. *PLoS ONE* **2019**, *14*, e0217301. [CrossRef]

32. Cappon, G.; Vettoretti, M.; Marturano, F.; Facchinetti, A.; Sparacino, G. A neural-network-based approach to personalize insulin bolus calculation using continuous glucose monitoring. *J. Diabetes Sci. Technol.* **2018**, *12*, 265–272. [CrossRef]

33. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.

34. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef]

35. Polydoros, A.S.; Nalpantidis, L. Survey of model-based reinforcement learning: Applications on robotics. *J. Intell. Robot. Syst.* **2017**, *86*, 153–173. [CrossRef]

36. Komorowski, M.; Celi, L.A.; Badawi, O.; Gordon, A.C.; Faisal, A.A. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat. Med.* **2018**, *24*, 1716–1720. [CrossRef]

37. Zhu, T.; Li, K.; Herrero, P.; Georgiou, P. Basal glucose control in type 1 diabetes using deep reinforcement learning: An in silico validation. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 1223–1232. [CrossRef]

38. Zhu, T.; Li, K.; Kuang, L.; Herrero, P.; Georgiou, P. An insulin bolus advisor for type 1 diabetes using deep reinforcement learning. *Sensors* **2020**, *20*, 5058. [CrossRef] [PubMed]

39. Kovatchev, B.P.; Breton, M.D.; Cobelli, C.; Dalla Man, C. Method, System and Computer Simulation Environment for Testing of Monitoring and Control Strategies in Diabetes United States. U.S. Patent 10,546,659, 28 January 2020.

40. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In Proceedings of the International Conference on Machine Learning, PMLR, Vienna, Austria, 25–31 July 2018; pp. 1861–1870.
41. Available online: https://github.com/Advanced-Learning-and-Automation/RL_for_MDI_Therapy_in-_T1D (accessed on 1 May 2023).