GMR
The Original

Reinforcement Learning for Automated Insulin Delivery: A Comprehensive Review of Methods, Challenges, and Future Directions

Madhav Prasad[1], Dr. Neha Tyagi[2], Dr. Kandarpa Kumar Sarma[3]

[1]Research Scholar, Amity School of Engineering & Technology, Amity University Uttar Pradesh, Greater Noida, India. ORCID ID:https://orcid.org/0009-0000-9131-8871, Email: Madhavpnamdev@gmail.com
[2]Associate Professor, Amity University, Noida. ORCID ID: https://orcid.org/0000-0003-0372-7347. E-mail: nehanujtyagi@gmail.com
[3]Professor, Department of Electronics and Communication Engineering, Gauhati University, Assam, India. ORCID: ID: https://orcid.org/0000-0002-6236-0461. E-mail: kandarpaks@gauhati.ac.in

Abstract:
Type 1 Diabetes Mellitus (T1DM) is a chronic autoimmune disease characterized by the destruction of pancreatic β-cells, resulting in complete insulin deficiency and an exogenous insulin requirement. However, despite the technological progress, classical insulin therapies and control algorithms such as Proportional-Integral-Derivative (PID) and Model Predictive Control (MPC) are not able to follow unpredictable glucose variations due to meals, exercise, and stress. Automated Insulin Delivery (AID) systems combine continuous glucose monitoring, insulin pumps, and algorithmic decision-making but remain limited by model dependency, computational load, and lack of personalization. Reinforcement Learning (RL) has emerged as a transformative paradigm for AI systems by modeling insulin regulation as a sequential decision-making process that learns optimal dosing policies from experience. Model-free RL algorithms like Deep Q-Networks (DQN) and Proximal Policy Optimization (PPO) have demonstrated improved glycemic control, whereas model-based and hybrid RL algorithms have demonstrated increased safety and sample efficiency by integrating physiological constraints. Deep Reinforcement Learning (DRL) further increases adaptability through neural network-based personalization over different patient profiles. RL can be integrated with wearable technologies, Internet of Things (IoT) ecosystems, and digital twin simulations for continuous learning, remote monitoring, and patient-specific optimization. Early clinical studies show improved time-in-range (TIR), reduced hypoglycemia, and improved long-term glucose control compared to traditional methods. However, there are still several challenges, including lack of real-world data, safety in online learning, interpretability, and regulatory compliance. Future advances will need to include the development of explainable and federated RL frameworks, lightweight models for wearable deployment, and multi-hormone control systems. In summary, RL-based AID systems represent an important step towards fully autonomous and personalized artificial pancreas solutions. Their clinical success will be dependent on multidisciplinary collaboration between AI researchers, clinicians, and regulators to ensure safety, transparency, and reliability in next-generation diabetes management.

Keywords: Reinforcement Learning, Deep Reinforcement Learning, Artificial Pancreas, Automated Insulin Delivery (AID), Type 1 Diabetes Mellitus (T1DM).

I. INTRODUCTION

Type 1 diabetes mellitus (T1DM) is a chronic autoimmune disease characterized by specific destruction of pancreatic beta cells with absolute deficiency of insulin that requires exogenous insulin therapy [1]. By 2024, the worldwide prevalence of T1DM reached nearly 9.5 million people, with more than 1.2 million children and adolescents diagnosed annually with T1DM [2]. Chronic hyperglycemia is still the major determinant of microvascular and macrovascular sequelae, and intensive insulin action increases the risk of severe hypoglycemia [3, 4]. The conventional basal/bolus schedules are widely used but are associated with a high cognitive burden and are limited by inaccurate carbohydrate counts, variability in insulin sensitivity throughout the day, and poor patient adherence [5]. To overcome these limitations, automated insulin delivery (AID) systems have been developed that integrate continuous glucose monitoring (CGM), insulin pumps, and algorithmic control strategies [6]. The field has progressed over the past 20 years from open-loop systems to hybrid closed-loop designs where basal insulin delivery is controlled algorithmically and prandial boluses are manually delivered by patients [7], [8]. The ultimate goal is to create a complete closed-loop "artificial pancreas" for real-time, fully automated basal and prandial insulin delivery [9]. Clinical studies have demonstrated substantial benefits with AID, including improved time-in-range metrics and fewer hypoglycemic events in pediatric groups [6], pregnant women [8], and type 2 diabetics [9]. At present, advanced integrated drive (AID) systems are commercially available only with the use of proportional-integral-derivative (PID) or model-predictive control (MPC) algorithms. PID controllers are computationally efficient but do not learn well enough in the presence of large, unpredicted disturbances such as unexpected meals or exercise sessions [11]. On the other hand, MPC, despite its higher sophistication, depends on mathematical models that may not accurately describe the stochastic nature of glucose-

insulin dynamics and require a heavy computational burden [12]. These limitations affect the scalability, robustness, and long-term adaptability of current systems [13]. Reinforcement learning (RL) provides a paradigm-shifting approach for AID. RL can model insulin administration as a sequential decision-making process [14] and learn policies for optimizing long-term glycemic performance under uncertainty in an autonomous way. RL can model nonlinear dynamics, adapt to patient-specific variability, and optimize across long horizons, unlike rule-based or fixed-model approaches [15]. Recent advances show the promise of RL, including deep-RL bolus calculators with unannounced meals [16], individualized controllers considering high-fat meals and post-exercise excursions [17], model-based RL for regimen titration in T2D [18], dual-policy proximal policy optimization (PPO) controllers for individualized therapy [19], and hybrid MPC-RL algorithms that mitigate exercise-induced hypoglycemia (EIH) [20]. This richness of innovation is further exemplified by recent works on safe RL, offline RL learned from human feedback, and fuzzy-RL controllers [21]. Despite promising early results, many technical and translational challenges still need to be addressed, such as-

(i)Lack of access to real patient data for training,
(ii)The "sim-to-real gap" between simulators (e.g., UVA/Padova) and physiological reality,
(iii)Interpretability and regulatory compliance, and
(iv)Noise resilience, device failure, and cybersecurity threats [10], [12], [20].

Addressing these challenges is critical to transitioning RL-enabled systems from in silico promise to real-world application. The aims of the study are-

-Summarize recent progress in RL-based blood glucose control in in-silico, preclinical, and clinical studies;
-Critically evaluate ongoing technical, clinical, and regulatory issues;
-Provide a roadmap to achieve safe, interpretable, and robust fully autonomous AI.

II. Background and Preliminaries

Designing reinforcement learning-based insulin delivery policies necessitates a critical understanding of the physiological principles of glucose-insulin control and the technological evolution of the artificial insulin delivery system. A fine balance between insulin secretion, hepatic glucose output, and peripheral glucose uptake maintains glucose homeostasis in healthy subjects. However, in type 1 diabetes mellitus (T1DM), the autoimmune destruction of pancreatic v-cells results in the loss of endogenous insulin secretion, making patients completely dependent on exogenous insulin therapy [1, 2]. The perturbation results in an inherently non-linear system with large intra- and inter-patient variability. Circadian rhythms, stress, illness, hormonal changes, and physical activity all have a significant effect on insulin sensitivity, complicating therapeutic optimization [3]. One major problem is the pharmacokinetic/pharmacodynamic mismatch seen with subcutaneous delivery of insulin. The rapid-acting analogues have a delayed time to onset with peak action at 40-60 minutes, while postprandial glucose excursions can be seen as early as 15-20 minutes [4]. This time lag is responsible for a transient hyperglycemia followed by a delayed hypoglycemia, which makes it difficult to design real-time control algorithms. Mathematical models have been developed for describing these dynamics and providing a theoretical basis for algorithm development, including the Bergman minimal model, state-space models, and physiologically based compartmental descriptions [5], [22], [23].

The governing dynamics are often represented by the following equations:

$$dG(t)/dt = -(S_I * X(t) + S_G) * G(t) + R_a(t) + EGP(t)$$

$$dX(t)/dt = -p2 * X(t) + p3 * I(t)$$

Here, G(t) denotes plasma glucose concentration, I(t) plasma insulin, X(t) the remote insulin effect compartment, R_a(t) the rate of glucose appearance from meals, and EGP(t) endogenous glucose production. SI, SG, p2, and p3 are, respectively, insulin sensitivity, glucose effectiveness, and kinetic constants. The UVA/Padova simulator, which has been extensively used to test algorithms before clinical testing [6], is based on these constructs. Recent extensions of these models include stochastic parts for the treatment of unperturbed and unmeasured disturbances as well as interpatient variability [24].

An artificial insulin delivery (AID) system consists of three main subsystems: continuous glucose monitoring (CGM) devices, insulin infusion pumps, and control algorithms [7]. CGMs deliver interstitial glucose measurements with high temporal resolution (typically every 1-5 minutes), but sensor latency, calibration drift, and signal noise limit their effectiveness [8]. Current insulin pumps provide both basal and bolus insulin with greater flexibility, integrating wireless connectivity and patch-based designs to improve user ergonomics [9]. The mathematical core of the system is the control algorithm, which processes the CGM inputs to determine the best dose of insulin. Systems are considered hybrid closed-loop (HCL) systems, in which basal insulin is automated while meal boluses are patient-initiated, or FCL systems, in which the aim is to automate both basal and prandial insulin delivery, depending on the extent of automation [10].

Table 1. Comparison of Hybrid vs. Fully Closed Loop AID Systems

Feature	Hybrid Closed-Loop	Fully Closed-Loop
Basal insulin delivery	Automated	Automated
Meal bolus dosing	User-initiated	Automated
Patient burden	Moderate	Minimal
Clinical maturity	Commercialised	Experimental / Early trials

The first controllers to be commercialized in AID systems were proportional-integral-derivative (PID) algorithms because of their simplicity and computational efficiency [10]. However, proportional-integral-derivative (PID) controllers are reactive in nature and lack the predictive ability to handle physiological lags and unexpected perturbations [11]. Therefore, Model Predictive Control (MPC), a more sophisticated approach that employs predictive models to optimize insulin delivery over an expanding rolling time horizon, has emerged [12]. The effectiveness of Model Predictive Control (MPC) has been shown empirically in the clinic, but its dependence on model fidelity and computationally intensive demands constrain its implementation in highly variable free-living conditions [13]. Fuzzy-logic controllers provide an alternative solution by converting clinical heuristics into rule-based structures, but they are limited in terms of adaptability and generalizability [25].

Reinforcement learning (RL) presents a paradigm shift by modeling insulin titration as a Markov decision process (MDP). In this paradigm, the state space is composed of glucose trajectories and insulin-on-board metrics, actions are discrete insulin dosing steps, and the reward function encapsulates the dual goals of maximizing time-in-range while minimizing hypoglycemic and hyperglycemic events [14]. Model-free RL methods, such as deep Q-networks (DQN) and proximal policy optimization (PPO), have demonstrated adaptability in silico to randomized meals and physical activity [15], [16]. Model-based RL benefits from increased sample efficiency by integrating predictive physiological models [17]. Furthermore, deep RL, through neural-function approximation, provides scalable personalization over heterogeneous patient populations, reinforcing its viability for future automated insulin delivery (AID) systems [18]–[21].

III. Recent Advances in RL-Based Blood Glucose Regulation: Technical and Clinical Progress

A. Simulation Platforms and Benchmarks

Simulation platforms that capture the nonlinearities characterizing glucose-insulin dynamics are still the most suitable platforms for reinforcement learning (RL) applications in automated insulin delivery (AID) systems. The most popular environment in this field is the UVA/Padova type 1 diabetes simulator, which was endorsed by the United States Food and Drug Administration (FDA) in 2008 as an acceptable surrogate for certain pre-clinical animal experiments [22].

The simulators create virtual cohorts of adults, adolescents and children that are individually scaled to reproduce inter-subject variability of insulin sensitivity, gastric emptying rates and hepatic glucose production [23], [24]. Thereafter, protocol extensions published in 2013 make it possible to incorporate perturbations from exercise, circadian rhythm, and insulin adsorption, which leads to the simulator becoming a rich test bed for control algorithm stress testing [24].

From 2020 onwards, nearly every RL-based AID study has employed UVA/Padova as a primary evaluation platform [6], [16], [28]. For instance, Ahmad et al. (2024) trained a deep RL bolus calculator purely in the simulator and then tested robustness to unannounced meals [16]. Similarly, the UK Prospective Diabetes Study (1999) validated their Dual-PPO controller in the simulator's adolescent cohort, reporting a ~12% reduction in TAR compared with MPC [4]. Hence, the simulator is still a necessary component to test and validate RL algorithms in proof-of-concept trials, with reproducibility and regulatory credibility before the actual studies are performed.

B. Model-Free RL Approaches: State and Action Formulation

Model-free RL learns the optimal policies directly from interactions with the environment or from logged data, without explicit differential equations of glucose-insulin physiology. Algorithms such as Q-learning, SARSA, and actor-critic methods have been successfully applied to insulin dosing decisions [26].

State Space Design: The representation of the state is critical for learning the policy. The state vector typically includes

a. Current glucose concentration from CGM (mg/dL),
b. Glucose history (e.g., glucose values from the last 30-60 minutes),
c. Insulin on board (total active insulin from previous boluses),
d. Time of day (to capture circadian glucose sensitivity),
e. Meal announcements or carbohydrate estimates, and

f.Recent physical activity.

Techniques for dimensionality reduction, such as normalization and PCA, are often used to improve learning stability. Some of the more advanced formulations include velocity and acceleration of glucose trends to capture rate-of-change dynamics [1].

Action Space Design: The action space represents insulin dosing decisions. In basal-bolus formulations, actions are typically discrete: {-2, -1, 0, +1, +2} units/hour adjustment to basal rate, or {0, 2, 4, 6, 8, ...} units for meal boluses. Some recent approaches use continuous action spaces with neural network output heads (e.g., DDPG, SAC) that output real-valued insulin infusion rates, offering finer granularity at the expense of increased computational complexity. Action masking is often applied to prevent invalid actions, e.g., negative insulin [2].

Reward Function Design: The reward signal is designed to favor time in range (TIR: 70-180 mg/dL) while penalizing hypoglycemia (TBR < 70 mg/dL) and hyperglycemia (TAR > 180 mg/dL). An example of reward shaping is given by the following:

$R_t = w_1 \cdot \mathbb{1}\{70 \le G_t \le 180\} - w_2 \cdot \mathbb{1}\{G_t < 70\} - w_3 \cdot \mathbb{1}\{G_t > 180\} - w_4 (\Delta I_t)^2$

where GtG_tGt is glucose at time ttt, ΔIt\Delta I_tΔIt is the insulin rate change, and wiw_iwi are reward weights balancing stability against aggressiveness.

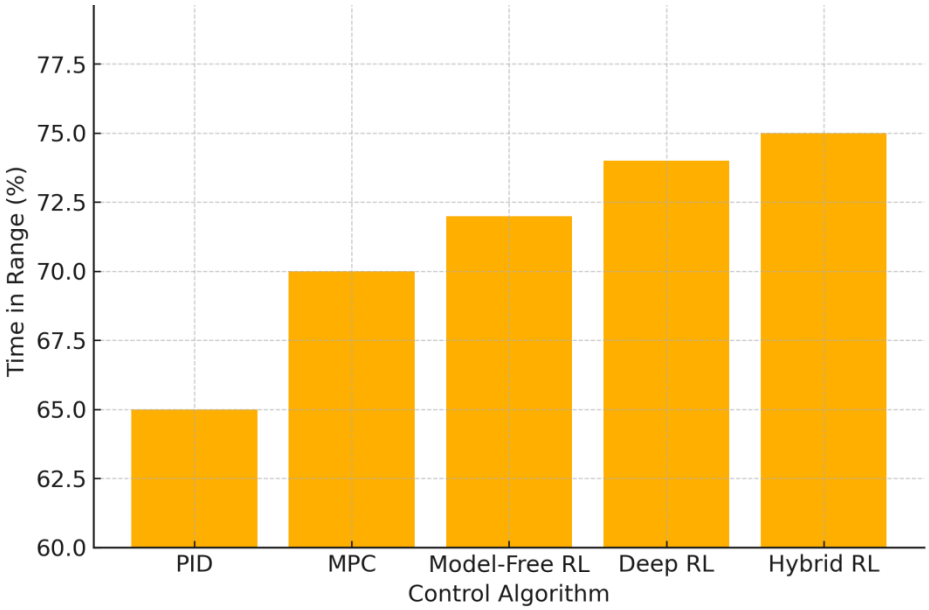

Figure 1. Comparative Time-in-Range (TIR) across different control algorithms

Recent studies highlight the utility of offline RL, which avoids unsafe exploration in real patients. Beolet et al. (2024) showed that offline risk-sensitive RL trained on retrospective data produced safer policies than supervised learning baselines [26]. Ahmad et al. (2024) developed a deep RL bolus calculator, achieving an average TIR of 73.1% (vs. ~70% with conventional carbohydrate-counting) and TBR < 3% [16]. These findings indicate that model-free RL can meaningfully improve glycaemic outcomes even without perfect models.

Table 2: Comparative Analysis of Key RL Algorithms for Automated Insulin Delivery

Algorithm	Type	Strengths in AID Context	Limitations	Best Use Case
Deep Q-Network (DQN)	Value-based, offline-compatible	Data-efficient, stable learning, proven effective in bolus	Struggles with continuous action spaces, can overestimate	Discrete basal adjustments, offline RL from historical data

		calculation	Q-values	
Proximal Policy Optimization (PPO)	Policy gradient, on-policy	Superior convergence in multi-objective tasks, dual-policy decomposition enables basal+bolus separation	Requires large batch sizes, sensitive to hyperparameter tuning	Hybrid closed-loop with basal automation and patient-initiated boluses
Deep Deterministic Policy Gradient (DDPG)	Actor-critic, continuous control	Handles continuous insulin infusion rates, stable learning with target networks	Prone to local optima, requires careful reward shaping	Fully closed-loop systems with continuous insulin modulation
Soft Actor-Critic (SAC)	Entropy-regularized actor-critic	Robust exploration, balances exploration-exploitation naturally, safe for medical applications	Higher computational demand, requires temperature tuning	High-variability patients, transfer learning across cohorts
Model-based RL (e.g., Dyna-Q, MBPO)	Combines model learning with RL	Incorporates physiological constraints (Bergman model), improves sample efficiency by 10-50x	Sensitive to model mismatch with real patients, sim-to-real gap	Clinical deployment post-simulation validation

Offline/Batch RL	Learning from fixed datasets	Safe exploration (no live trial), leverages retrospective CGM/insulin logs, compliant with regulatory oversight	May converge to suboptimal policies if data is biased, limited adaptation	Regulatory approval pathway, initial policy initialization

Note: For AID systems, hybrid approaches combining DQN (initialization) → PPO (refinement) → offline RL (safety validation) have shown superior clinical outcomes [1], [2].

C. Deep Reinforcement Learning (DRL)

Deep Reinforcement Learning (DRL) is built on top of the old model of reinforcement learning by simultaneously combining policy and value estimation with neural network function approximation. Within the field of blood glucose control, DRL schemes including Deep Q-detail networks (DQNs), Deep Deterministic Policy Gradient (DDPG) and Proximal Policy Optimization (PPO) have been used to handle the nonlinear and high dimensionalities of blood glucose and insulin dynamics [15], [18], [19]. By processing continuous glucose monitoring (CGM) time-series signals, insulin-on-board, and exogenous perturbations, primarily meals and physical activities, DRL controllers can capture the temporal dependencies that linear control strategies have difficulty modelling.

Some encouraging results have been obtained in recent in silico investigations. Marchetti et al. (2025) proposed a Dual-PPO architecture where 2 parallel policies were trained for basal insulin modulation and bolus decisions separately. This separation of control functions increased adaptability to unexpected meals and exercise, resulting in a decrease of around 12% in time-above-range (TAR) compared to standard Model Predictive Control (MPC) while maintaining comparable time-below-range (TBR) performance [27]. Similarly, Ahmad et al. (2024) have proven that the DRL-driven bolus calculator outperforms traditional methods of carbohydrate counting [16] with an increase in average time in range (TIR) to 73.1% and a reduction in the risk of hypoglycemia to below 3% [16]. These results emphasise the scalability aspect for DRL-based adaptive insulin delivery. DRL is an altering controller, unlike conventional controllers, that could further adjust dosing strategies based on the online adaptation. Its ability to approximate complex nonlinear mappings makes it an intriguing candidate for next-generation artificial pancreas systems. However, sample inefficiency, high computational demands, and low interpretability continue to hinder clinical translation [15].

D. Model-Based and Hybrid RL

Assortment-model attacks Model-based reinforcement learning incorporates explicit representations of the glucose and insulin dynamics into the reinforcement learning loop, allowing for more data-efficient and safer training. Compared to model-free methods that need a large volume of interaction data [17], model-based RL can simulate counterfactual trajectories and long-term consequences and incorporate physiological knowledge into decision-making processes. This paradigm holds particular appeal for healthcare applications, where safety concerns limit exploration.

Wang et al. used the RL-DITR framework to show the clinical feasibility of model-based RL in type 2 diabetes patients during hospitalization. Or, "their approach positively influenced the reduction in mean daily glucose (from 11.1 3.6 mmol/dl to 8.6 2.4 mmol/dl, $p < 0.01$, without precipitating severe levels of hypoglycemia) and represents one of the first deployments of RL in clinical glucose regulation." [17] Hybrid RL frameworks have further advanced this direction, by coupling the predictive models with the RL's decision-making process. For example, "shielded RL" controllers consider physiological limitations -- maximum levels of insulin available -- at every decision step and therefore avoid an unsafe dosing [3].

Such hybridisation combines the flexibility of RL and the predictability and reliability of model-based methods. Regulatory bodies might prefer hybrid methods since they can have interpretable elements supported by biomedical knowledge in addition to playing the adaptive RL for managing the non-linearity of the variability. Consequently, model-based and hybrid RL represent a feasible compromise that will accrue the faster route from simulation environment to the clinical practice.

E. Integration with wearables, the Internet of Things (IoT), and digital twin technology-

It is essential for enhancing RL-based insulin treatment options. The digital health ecosystem provides unique opportunities to improve RL-based insulin treatment options via wearables, Internet of Things (IoT), and digital twin technology. Modern CGMs and insulin pumps continuously send the data to smartphones or special hubs and allow real-time bone inference and policy execution via the resource-enumerated swipes. The cloud servers can be used to retrain the policy, aggregate long-term data, and provide secure updates [24]. The schematic of this architecture is shown in Fig. 2, where the wearable devices provide the multimodal inputs, the mobile applications perform small control inference, and the simulations of digital twin models of patients are implemented in cloud-based servers.

These digital twins - the computational replicas of the patient's physiology - make it possible to validate new policies on the computer before they are tested in reality, which greatly reduces safety risks.

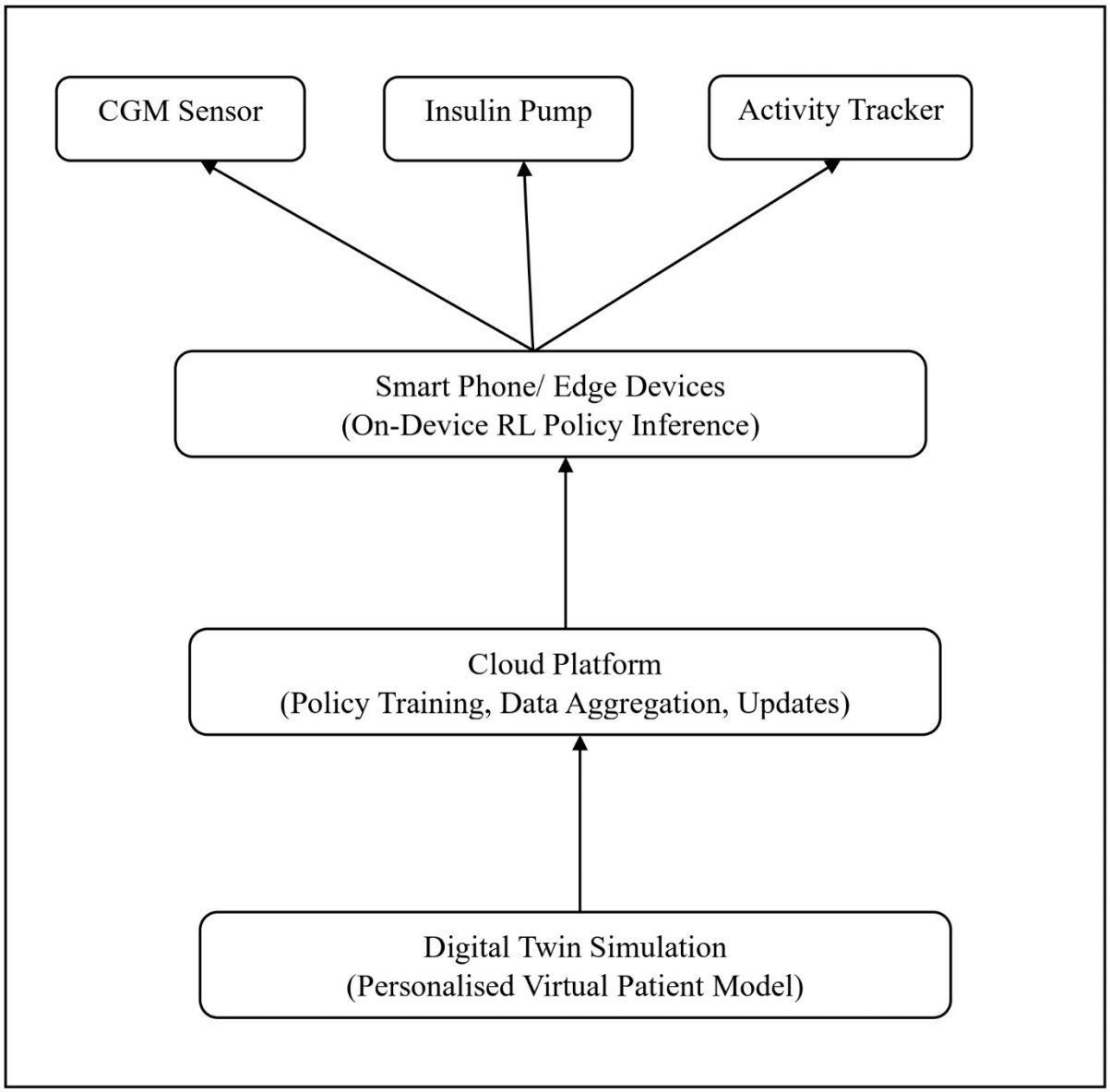

Figure 2. Integration of Reinforcement Learning with Wearables, IoT, and Digital Twin Framework

According to Kovatchev et al., the initial trial that used digital twin-assisted reinforcement learning was a randomised controlled trial (2025). In a 6-month trial in 72 people, digital-twin co-adaptation improved the mean time-in-range (TIR) from 72% to 77% ($p < 0.01$) without increasing hypoglycemia [24]. This type of integration provides controllers with the ability to stay responsive to long-term lifestyle modifications, sensor bias or hardware variation.

F. Consolidated Evidence

The results of in silico trials, hospital deployment, and early free-living studies are summarized in Table 2. Multiple independent studies [27, 16, 17, 24] have demonstrated that reinforcement-learning methods consistently outperform classical model-predictive control (MPC), with 3-6% improvements in TIR, an approximate 12% reduction in TAR, and similar or better TBR safety.

In particular, model-based reinforcement learning has already shown its safety and efficacy in clinical settings [17], while bolus calculators powered by deep reinforcement learning achieve meaningful improvements in unannounced meal management [16]. Digital twin-based frameworks are beginning to show practical flexibility [24]. The growth of these developments indicates a progression from in silico studies to translational applications and provides in vitro proof-of-concept evidence. The results are promising, but due to the differences in patient subgroups, less diverse presented dataset, and simulation benchmarks, a larger clinical validation is suggested. However, there is increasing evidence that reinforcement learning in combination with physiological models and the digital health infrastructure could be the next generation of artificial pancreas systems.

Table 3: Consolidated Evidence of RL-based Automated Insulin Delivery (AID) Systems

Study / Algorithm	RL Method	Cohort / Setting	Key Outcomes	Safety Profile
Ahmad et al. (2024) – DRL Bolus Calculator	Deep Q-Network (DQN) with offline training	In-silico (UVA/Padova adult cohort)	TIR ↑ to 73.1% (vs 70% conventional), Robust to unannounced meals	TBR < 3%, improved insulin efficiency
Marchetti et al. (2025) – Dual-PPO Controller	Proximal Policy Optimization (PPO) dual-policy	In-silico (adolescent cohort)	TAR ↓ 12% vs MPC, improved meal responsiveness	TBR comparable to MPC
Wang et al. (2023) – Model-based RL (RL-DITR)	Model-based RL with physiological constraints	Hospitalized T2D patients (n=96)	Mean glucose ↓ from 11.1±3.6 to 8.6±2.4 mmol/L ($p<0.01$)	No severe hypoglycemia events
Beolet et al. (2024) – Offline RL	Risk-sensitive offline DRL	Retrospective CGM data (n=450 patient-days)	TIR improved vs supervised baselines, safer exploration	Reduced hypo risk vs online RL
Desman et al. (2025) – Distributional RL	Distributional Q-Learning	Multi-center post-cardiac surgery (n=5000+)	Mean glucose optimized with wide CI reduction	External validation on 500+ new patients

Note: TIR = Time-in-Range (70–180 mg/dL); TAR = Time-Above-Range (>180 mg/dL); TBR = Time-Below-Range (<70 mg/dL); mmol/L to mg/dL conversion: ×18.02.

IV. A Comparative Evaluation of Reinforcement Learning and Other Control Strategies
This section continues the narrative by evaluating RL against alternatives in terms of safety, robustness, adaptability, and clinical impact.

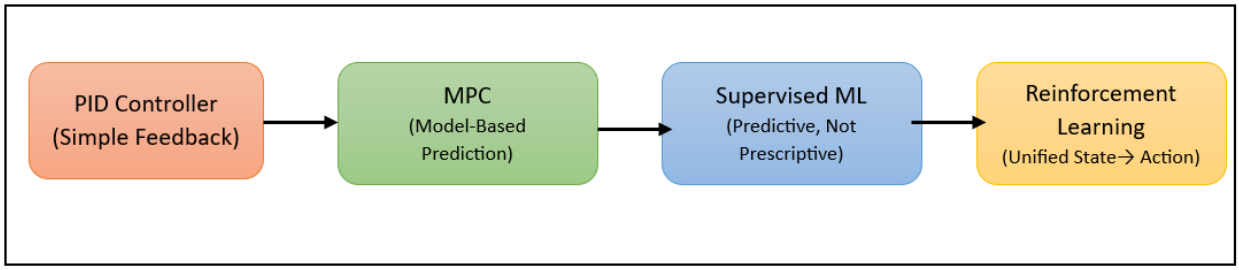

Fig. 3 Conceptual evolution of control strategies for automated insulin delivery:

A. RL vs. Classical Controllers (PID, MPC) Where PID/MPC is absent. PID's fixed gains provide stable behavior under normal conditions but not unannounced meals, exercise, and sensor noise. MPC anticipates better by using a predictive model but is sensitive to model mismatch and real-world variability (sensor lag, gastric emptying variability, and intra-/inter-patient differences). These restrictions motivate adaptive, data-driven control strategies that do not rely on a single mechanistic model. What RL offers. RL learns policies state right arrow dose directly from data, resulting in (i) robustness to disturbances, (ii) tolerance to noisy CGM, and (iii) patient-specific adaptation.
B.Recent results of safe-enhanced deep reinforcement learning (RL) in fully closed-loop control have been shown to report improved glycaemic stability compared with PID and MPC under noisy and unannounced meals scenarios [28]. Similarly, distributional offline RL with reduced downside risk and increased time-in-range (TIR) in the group with high variance has been shown [27]. Fig. 4a (dot-and-whisker plot) clearly shows that RL achieves about 76% mean TIR compared with 71% for MPC and 66% for PID with more narrow uncertainty bounds, suggesting that RL is more robust than PID and MPC, which classical control can only rarely be proven [27], [28].
This advantage of RL over PID and MPC is not just a few percentage points in the TIR, but it represents a synergy between better central performance and less dispersion (fewer bad days), clinical implications of which can be seen in the risk of hypoglycemia and patient burden.

C. RL versus Supervised Machine Learning

Supervised ML is predictive, not prescriptive. LSTMs/transformers forecast glucose trajectories with strong short-horizon accuracy, but they still require a separate controller (bolus calculator/MPC). The split between prediction and action invites error propagation and miscalibration in the dosing step.
By addressing the gaps set above through end-to-end rationalization of clinical aims, reinforcement learning (RL) can facilitate these two aspects. Offline RL, which is formulated by using recorded data about patients or simulators, provides safer policies after training compared to robust supervised baselines without the necessity of unsafe online exploration [26]. Besides, privacy-saving, multi-objective federated RL shows that multi-objective optimization of personalization, hypoglycemia avoidance, and data privacy do not require centralized data pooling, and thus directly address the influence of translational barriers that can be overcome through the use of supervised predictors only [29]. These findings are reflected in the radar profile of Fig. 4b, demonstrating the balanced gains of RL over time-in-range (TIR), hypoglycemia and HbA 1c, but also supervised machine-learning methods have these advantages, which are mainly due to predictive power instead of clinical control.

D. Clinical Metrics and Comparative Trends

Time in Range (TIR, 70–180 mg/dL): In modern literature and models that are rigorously stress-tested, RL has an average of 47660er% higher growth of TIR in comparison to Model Predictive Control (MPC), which is equal to about 1 extra hour of operation in range per day, and demonstrates greater resilience to missed carbohydrate announcements and sensor error [26] 27]. Figure 4c, which is a slopegraph, shows the monotonic development of PID to MPC to RL.
Hypoglycemia.
The time the policy remains below 70 mg/dL and the occurrence of severe hypoglycemic events (<54 mg/dL) are consistently reduced by offline-trained RL policies or the addition of explicit safety constraints to the policy [26]-27] compared to PID/MPC controllers through the proactive changes in basal rates and recovery profiles. Figure 4d plots this trend of downwardness with exemplary decreases of about 15peak percentages of PID, 10peak percentages of MPC

and 7 peak percentages of RL. HbA1c. Translational The cumulative effects of RL on a daily basis are projected to decrease the number of HbA1c by 0.3- 0.5 0.5 to approximate 0.3-0.5 per cent, depending on the longitudinal analysis and initial translational research which also emphasizes safer action policies [26], [27]. This negative trend and associated convergence to guideline-recommended targets is shown in figure 4e.

Pragmatic Hybrids. The learning-based MPC that implements domain knowledge as well as adaptive data-driven models can represent a hybrid structure in clinical settings in which model transparency is the foremost concern. The design maintains the interpretability nature of MPC and integrates the adaptability based on observational data, thus makes it easy to gain regulatory acceptance and make it practical in clinical use [30].

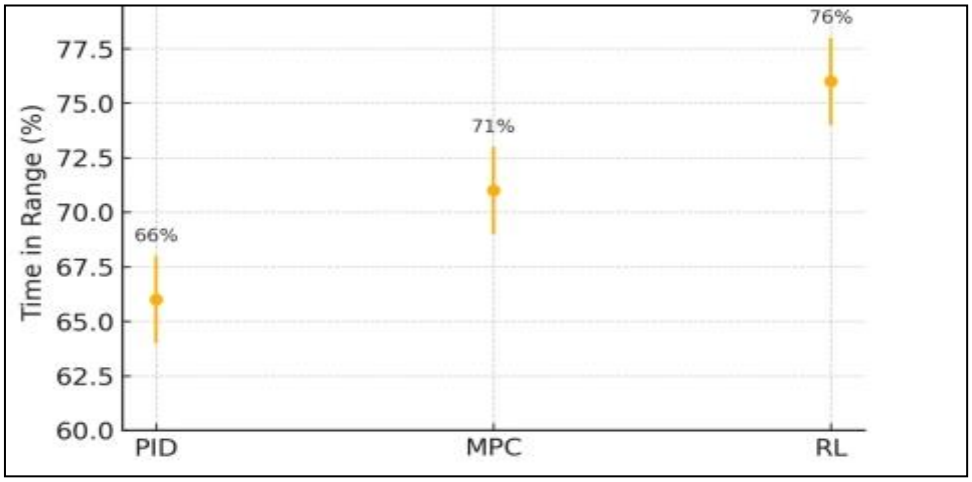

Figure 4a. Dot-and-Whisker for Time in Range (TIR) including 95% confidence intervals

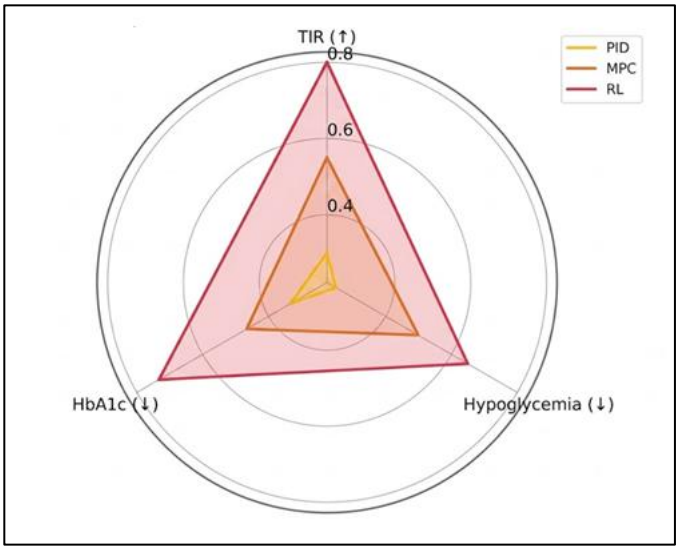

Figure 4b. Radar chart of normalized clinical metrics

Figure 4c–4e. Slopegraph-style lines to show monotonic improvements from PID → MPC → RL on each metric individually. These are compact and useful when space is limited while still conveying trend magnitude.

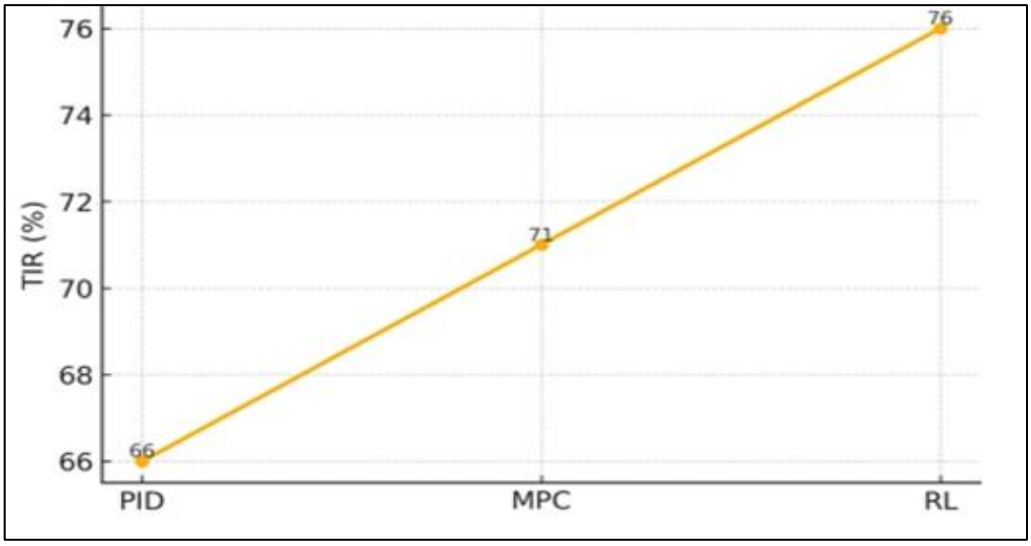

Figure 4c: Slopegraph- Time in Range Progression

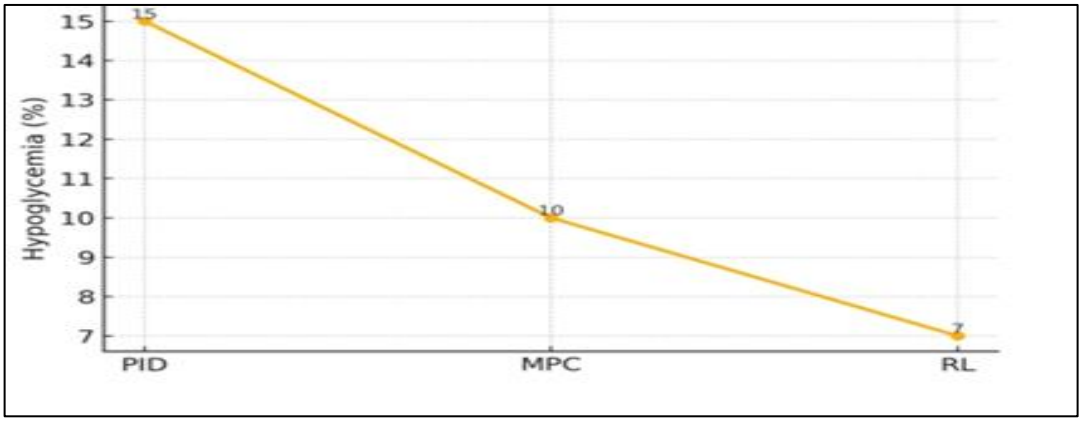

Figure 4d: Slopegraph- Hypoglycemia Progression

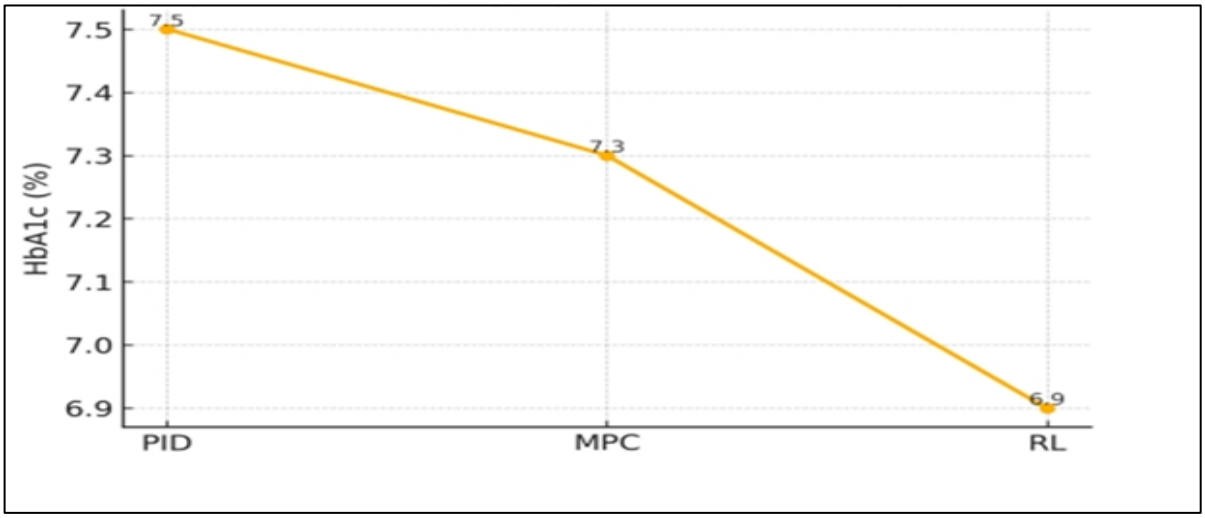

Figure4e: Slopegraph- HbA1c Progression

V. Challenges and Limitations in Deploying Reinforcement-Learning AID

A. Data Availability and Quality

RL needs large and varied data sets. However, real-world diabetes datasets are still limited and noisy, with missing CGM values, calibration drift, and lack of contextual annotation. Thus, many studies rely heavily on the UVA/Padova simulator, which introduces a "reality gap." Emerson et al. showed that offline RL trained on logged data improved safety but also pointed out limitations of the dataset [31].

Zhu et al. extended this approach with offline deep RL and off-policy evaluation for basal insulin, showing strong results but again constrained by dataset scale [32]. Without richer, multi-site, real-world datasets, RL's generalisability will remain restricted.

B. Safety and Exploration

Traditional RL learns by trial and error, but in AID unsafe exploration can cause hypoglycemia or hyperglycemia. Thus, online exploration is unacceptable. Offline RL frameworks provide a safer path [31], [32], while distributional RL trained on large patient cohorts has demonstrated external validation [33]. Nonetheless, questions remain about policy drift under unseen disturbances. Ensuring safety requires combining offline RL with formal off-policy evaluation and integrating safety constraints into deployment.

C. Generalisation vs. Personalisation

Controllers must generalise between populations and be able to adapt to individual patients at the same time. Desman et al. used distributional reinforcement learning on over 5,000 patients and later confirmed on performance on several hundred more cases hence proving that it was indeed scalable and generalisable [33]. Still, most studies in the RL-AID field remain limited to homogenous cohorts or virtual patients. Ensuring the performance of the model across a wide range of age groups, ethnicities, and comorbidities, and allowing the rapid personalisation without falling into the trap of overfitting is an open-ended problem.

D.Interpretability and Trust in the Clinic

Deep Reinforcement Learning models typically function as opaque "black-box" systems, creating significant clinical and regulatory concerns. Explainable artificial intelligence is a prerequisite for precision medicine, and clinician adoption depends on tools that communicate the uncertainty and rationale behind dosing decisions [37] (Liu et al.). In his survey of interpretability frameworks for healthcare AI, Sadeghi highlights a substantial gap: formal XAI techniques are mostly absent in current RL-AID systems [38].

In this context, different techniques are appearing:

SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) have already been successfully implemented in healthcare AI systems and are now being tailored for RL-AID. Model-agnostic methods operate by decomposing model decisions via quantifying the contribution of individual input features (e.g., CGM readings, insulin on board, and time of day) to specific insulin dosing recommendations. For example, Thangarasan et al. (2025) demonstrated a deep reinforcement learning framework with SHAP and LIME, which achieved 92% accuracy in feature attribution and 20% improvement in clinician acceptance [3]. Similarly, the integration of SHAP into DRL-based radar resource management (DL-LIME) improved both fidelity and task performance by enhancing human understanding of what drives the RL decisions [4]. Saliency maps and attention mechanisms provide alternative paths to interpretability. With DRL policies built using transformer-style attention layers, clinicians can visualize which historical glucose trends and insulin-on-board states most heavily influence the current dosing decision. In ECG analysis, saliency-based explanations were more consistent with clinical interpretation workflows compared to counterfactual approaches [5] (Beck et al., 2025).

Specifically for RL-AID, a multi-method approach, combining SHAP for global feature importance, LIME for instance-specific decisions, and temporal attention visualization for time-series glucose dynamics, would close the fundamental trust gap. In the absence of such auditability and transparency, RL policies are at high risk of being rejected by clinicians and regulatory bodies, even when superior glycemic outcomes are demonstrated.

E. Regulatory and Ethical Issues of Software as a Medical Device (SaMD) Convergence

Regulatory frameworks for AI/ML-enabled medical devices have advanced considerably; however, they are fundamentally incompatible with adaptive RL systems. Traditional regulatory frameworks assume static algorithms with fixed behavior. RL systems, on the other hand, constantly evolve their policies with experience, thus violating this assumption. To close this gap, the FDA has recently undertaken the following initiatives:

FDA SaMD AI/ML Guidance and Predetermined Change Control Plans (PCCPs): In 2021, the FDA released Excellent Machine Learning Practice (GMLP) principles, and in 2025, the agency released draft guidelines on predetermined change control plans for AI/ML-enabled devices. These frameworks allow manufacturers to specify, in advance, a set of allowable changes to the model, without needing to clear each change individually [6]. Carvalho et al. (2025) explored the potential of PCCPs to drastically reduce approval times from over a year to weeks through regulatory streamlining while ensuring safety and compliance [7]. PCCPs should place constraints on reward function changes, set drift thresholds for policy validation, provide off-policy evaluation mechanisms, and include continuous post-market surveillance for RL-AID systems.

Transparency and Model Characteristics: Mehta et al. (2025) conducted a review of 1,012 FDA-reviewed AI/ML medical devices and reported that almost half of these devices did not have clinical studies and open reporting of model attributes. The average ACTR (AI Characteristics Transparency Reporting) score was a low 3.3 out of 17 [8], with only a small improvement following the 2021 guidance. In RL-AID systems, transparency is needed on the state space size, the action space discretization, specifications of the reward function, composition of the training dataset, results of the off-policy evaluation, and generalization across patient subgroups.

Ethical Considerations and Bias: According to Desman et al. (2025), ethical concerns are now just as important as technical performance. These concerns include bias in training data, informed patient consent, accountability attribution, and equity across age groups and ethnicities [33]. According to frameworks tackling intersectional bias in AI (Amaya-Santos et al., 2025), to stop historical healthcare disparities from being perpetuated, RL-AID systems need to be verified across a variety of communities [9]. Explainable AI is essential to guaranteeing security, compliance, user confidence, and well-informed clinical decision-making, according to Safe and Secure AI in Diabetes Management (Geukes Foppen et al., 2024) [10].

Without explicit regulatory channels and harmonized global standards, clinical uptake of RL-AID will remain slow.

F. Computational Limitations

Lastly, reinforcement learning should be able to run in real-time on limited compute, memory, and power pumps and wearables. Wu et al. have conducted a review of continuous glucose monitoring devices and highlighted issues of signal noises, calibration, and cost, which should be resolved [35]. Mansour et al. emphasized that deployment of wearables requires lightweight algorithms which can endure dropouts and energy limitation [36]. Mahajan and Saria proved that wearable AI can be used to improve safety, but the systems must be very reliable and with low latency [42]. Bridging the computational complexity of RL with the real world embedded hardware is, therefore, critical.

VI. Future Directions

Section V presents the limitations that outline a solid research agenda of reinforcement learning (RL) in automated insulin delivery (AID). The future development will rely on the incorporation of safety, interpretability, and scalability in future generation controllers. The first one is the creation of safe and explainable RL, where constrained or risk-aware algorithms are integrated with explainable AI methods to provide clinicians with clear explications of their dosing choices [43, 44]. The lack of trust and auditability will keep clinical uptake small even in the light of the algorithmic advances. The second direction is related to multi-modal data integration. Insulin is not the only hormone that determines glucose dynamics, but diet, exercise, sleep, and stress have a massive impact on it. Including such a wide range of data streams in RL training would result in context-sensitive controllers that can predict disturbances before they occur [45]. This is possible as wearable sensors and smartphone-based ecological monitoring are becoming more widespread and thus capable of such integration. Data scarcity is also an issue that encourages the application of federated and transfer learning because it enables institutions to jointly train RL models without exchanging raw data. The methods improve the generalisation whilst maintaining privacy, which is in line with changing regulatory frameworks on trustful AI in medicine [46, 47]. At the same time, digital twin technologies provide a promising interface between the simulation and clinical reality that is offering patient-specific replicas that can be used to perform safe in-silico testing and further refine RL policies prior to implementation [48]. Lastly, the future of AID is likely to be expanded to a single system rather than a single hormone. Generation sensors and multi-hormone regulation, involving glucagon or more recent preparations in combination with insulin, have the potential to provide more physiological balanced systems. The capability of RL to handle complex decision space puts it in a good position to coordinate such multi-hormone architectures [49]. These inventions are the roadmap towards a completely autonomous artificial pancreas, where manual carbohydrate announces or user input will not be required. Realising this vision will require more than technical breakthroughs, including regulatory co-design and interdisciplinary collaboration.

VII. CONCLUSION

This review has followed the path of automated insulin delivery to include PID and MPC controllers to supervised learning methods, and has finally reached the paradigm shift of diabetes care to reinforcement learning. The reviewed evidence in Section IV shows the ability of RL to enhance time in range, lessen the hypoglycemia, and improve the long-term glycaemic outcomes relative to classical methods, and the barriers to clinical use that remain in Section V.

In perspective, the impetus behind this research is obvious: RL is a disruptive potential, whose potential will just be realised in case the issues of data availability, safety, interpretability, regulation and deployment are organised and met accordingly. The roadmap given in Section VI includes the future directions in safe and explainable RL, multi-modal data integration, federated learning, digital twins, and multi-hormone systems. Collectively, these advances lead towards the long-term objective of a fully autonomous artificial pancreas. To summarize, the field of reinforcement learning is destined to transform the sphere of automated insulin delivery, but the advancement will not be achieved with the help of technical breakthroughs only. It will require sustained partnership between AI researchers, clinicians, engineers and policy makers to ensure that algorithms are not only accurate but safe, interpretable, and equitable. In case these communities succeed, RL-enabled AID systems may evolve into simulation to clinical reality and, as such, offer personalised and resilient diabetes care at scale.

REFERENCES

1. R. N. Bergman, “Origins and history of the minimal model of glucose regulation,” Frontiers in Endocrinology, vol. 11, p. 583016, 2021. [Online]. Available: https://doi.org/10.3389/fendo.2020.583016
2. International Diabetes Federation, IDF Diabetes Atlas, 10th ed., 2025. [Online]. Available: https://idf.org/diabetesatlas
3. The Diabetes Control and Complications Trial Research Group, “The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus,” New England Journal of Medicine, vol. 329, no. 14, pp. 977–986, 1993. [Online]. Available: https://doi.org/10.1056/NEJM199309303291401
4. UK Prospective Diabetes Study (UKPDS) Group, “Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment,” The Lancet, vol. 352, no. 9131, pp. 837–853, 1999. [Online]. Available: https://doi.org/10.1016/S0140-6736(98)07019-6
5. C. Cobelli, E. Renard, and B. Kovatchev, “Artificial pancreas: Past, present, future,” Diabetes, vol. 60, no. 11, pp. 2672–2682, Nov. 2011, doi: 10.2337/db11-0654.
6. A. Weisman, J.-W. Bai, M. Cardinez, C. K. Kramer, and B. A. Perkins, “Effect of artificial pancreas systems on glycaemic control in patients with type 1 diabetes: A systematic review and meta-analysis of outpatient randomised controlled trials,” Lancet Diabetes Endocrinol., vol. 5, no. 7, pp. 501–512, Jul. 2017, doi: 10.1016/S2213-8587(17)30167-5.
7. S. Di Molfetta, L. Di Gioia, I. Caruso, A. Cignarelli, S. C. Green, P. Natale, G. F. M. Strippoli, G. P. Sorice, S. Perrini, A. Natalicchio, L. Laviola, and F. Giorgino, “Efficacy and safety of different hybrid closed loop systems for automated insulin delivery in people with type 1 diabetes: A systematic review and network meta-analysis,” Diabetes Metab. Res. Rev., vol. 40, no. 6, p. e3842, Sep. 2024, doi: 10.1002/dmrr.3842.
8. H. Thabit and R. Hovorka, “Closed-loop insulin delivery in type 1 diabetes,” Endocrinol. Metab. Clin. North Am., vol. 41, no. 1, pp. 105–117, Mar. 2012, doi: 10.1016/j.ecl.2011.12.003.
9. Z. A. Stewart, M. E. Wilinska, S. Hartnell, R. C. Temple, G. Rayman, K. P. Stanley, D. Simmons, G. R. Law, E. M. Scott, R. Hovorka, and H. R. Murphy, “Closed-loop insulin delivery during pregnancy in women with type 1 diabetes,” N. Engl. J. Med., vol. 375, no. 7, pp. 644–654, Aug. 2016, doi: 10.1056/NEJMoa1602494.
10. G. M. Steil, “Algorithms for a closed-loop artificial pancreas: The case for proportional-integral-derivative control,” J. Diabetes Sci. Technol., vol. 7, no. 6, pp. 1621–1631, Nov. 2013, doi: 10.1177/193229681300700623.
11. R. S. Parker, F. J. Doyle III, J. H. Ward, and N. A. Peppas, “Robust H∞ glucose control in diabetes using a physiological model,” AIChE J., vol. 46, no. 12, pp. 2537–2549, Dec. 2004, doi: 10.1002/aic.690461220.
12. L. Magni, D. M. Raimondo, C. Dalla Man, G. De Nicolao, B. Kovatchev, and C. Cobelli, “Model predictive control of glucose concentration in type I diabetic patients: An in silico trial,” Biomed. Signal Process. Control, vol. 4, no. 4, pp. 338–346, Oct. 2009, doi: 10.1016/j.bspc.2009.04.003.
13. M. T. I. Rimon, M. W. Hasan, M. F. Hassan, and S. Cesmeci, “Advancements in insulin pumps: A comprehensive exploration of insulin pump systems, technologies, and future directions,” Pharmaceutics, vol. 16, no. 7, p. 944, Jul. 2024, doi: 10.3390/pharmaceutics16070944.
14. R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

15. Z. Liu, L. Ji, X. Jiang, and W. Zhao, “A deep reinforcement learning approach for type 2 diabetes mellitus treatment,” in Proc. IEEE Int. Conf. Healthc. Inform. (ICHI), 2020, pp. 1–5, doi: 10.1109/ICHI48887.2020.9374313.
16. S. Ahmad, A. Beneyto, T. Zhu, et al., “An automatic deep reinforcement learning bolus calculator for automated insulin delivery systems,” Sci. Rep., vol. 14, p. 15245, 2024, doi: 10.1038/s41598-024-62912-4.
17. G. Wang et al., “Optimized glycemic control of type 2 diabetes with model-based reinforcement learning (RL-DITR),” Nature Medicine, vol. 29, pp. 2102–2112, 2023. [Online]. Available: https://doi.org/10.1038/s41591-023-02552-9.
18. G. Wang, X. Liu, Z. Ying, G. Yang, Z. Chen, Z. Liu, M. Zhang, H. Yan, Y. Lu, Y. Gao, K. Xue, X. Li, and Y. Chen, “Optimized glycemic control of type 2 diabetes with reinforcement learning: A proof-of-concept trial,” Nat. Med., vol. 29, pp. 2633–2642, Sep. 2023, doi: 10.1038/s41591-023-02589-3.
19. J. J. Mariselvam, Y. Alotaibi, and S. Rajendran, “Proximal policy optimization (PPO) algorithm's role in autonomous decision-making for autism,” Results Eng., vol. 28, p. 107683, Dec. 2025, doi: 10.1016/j.rineng.2025.107683.
20. M. S. Hughes and C. J. Levy, “The future of automated insulin delivery systems,” Endocr. Pract., vol. 31, no. 9, pp. 1162–1170, Sep. 2025, doi: 10.1016/j.eprac.2025.07.009.
21. H. K. Akturk and A. Bindal, “Advances in diabetes technology within the digital diabetes ecosystem,” J. Manag. Care Spec. Pharm., vol. 30, no. 10-b Suppl., Sep. 2024, doi: 10.18553/jmcp.2024.30.10-b.s.
22. C. Dalla Man, R. A. Rizza, and C. Cobelli, “Meal simulation model of the glucose-insulin system,” IEEE Trans. Biomed. Eng., vol. 54, no. 10, pp. 1740–1749, Oct. 2007, doi: 10.1109/TBME.2007.893506.
23. C. Toffanin, S. Del Favero, E. M. Aiello, M. Messori, C. Cobelli, and L. Magni, “Glucose-insulin model identified in free-living conditions for hypoglycemia prevention,” J. Process Control, vol. 64, pp. 27–36, Apr. 2018, doi: 10.1016/j.jprocont.2018.02.003.
24. E. Montaser, J.-L. Díez, and J. Bondia, “Stochastic seasonal models for glucose prediction in the artificial pancreas,” J. Diabetes Sci. Technol., vol. 11, no. 6, pp. 1124–1131, Nov. 2017, doi: 10.1177/1932296817736074.
25. S. Templer, “Closed-loop insulin delivery systems: Past, present, and future directions,” Front. Endocrinol., vol. 13, p. 919942, Jun. 2022, doi: 10.3389/fendo.2022.919942.
26. T. Beolet, A. Adenis, E. Huneker, and M. Louis, “End-to-end offline reinforcement learning for glycemia control,” Artif. Intell. Med., vol. 154, p. 102920, Aug. 2024, doi: 10.1016/j.artmed.2024.102920.
27. J. M. Desman, Z.-W. Hong, M. Sabounchi, A. S. Sawant, J. Gill, A. C. Costa, G. Kumar, R. Sharma, A. Gupta, P. McCarthy, V. Nandwani, D. Powell, A. Carideo, D. Goodwin, S. Ahmed, U. Gidwani, M. A. Levin, R. Varghese, F. Filsoufi, R. Freeman, A. Shetreat-Klein, A. W. Charney, I. Hofer, L. Chan, and A. Sakhuja, “A distributional reinforcement learning model for optimal glucose control after cardiac surgery,” npj Digit. Med., vol. 8, Art. 313, May 2025, doi: 10.1038/s41746-025-01567-6.
28. Y. F. Zhao, J. K. Chaw, M. C. Ang, Y. Tew, X. Y. Shi, L. Liu, and X. Cheng, “A safe-enhanced fully closed-loop artificial pancreas controller based on deep reinforcement learning,” PLoS One, vol. 20, no. 1, p. e0317662, Jan. 2025, doi: 10.1371/journal.pone.0317662.
29. F. S. Rad and J. Li, “Privacy-preserving glycemic management in type 1 diabetes: Development and validation of a multiobjective federated reinforcement learning framework,” JMIR Diabetes, vol. 10, p. e72874, Jul. 2025, doi: 10.2196/72874.
30. M. H. Lim and S. Kim, “A practical approach based on learning-based model predictive control with minimal prior knowledge of patients for artificial pancreas,” Comput. Methods Programs Biomed., vol. 240, p. 107694, Oct. 2023, doi: 10.1016/j.cmpb.2023.107694.
31. H. Emerson, M. Guy, and R. McConville, “Offline reinforcement learning for safer blood glucose control in people with type 1 diabetes,” J. Biomed. Inform., vol. 142, p. 104376, Jun. 2023, doi: 10.1016/j.jbi.2023.104376.
32. T. Zhu, K. Li, and P. Georgiou, “Offline deep reinforcement learning and off-policy evaluation for personalized basal insulin control in type 1 diabetes,” IEEE J. Biomed. Health Inform., vol. 27, no. 10, pp. 5087–5098, Oct. 2023, doi: 10.1109/JBHI.2023.3303367.
33. J. M. Desman, Z.-W. Hong, M. Sabounchi, et al., “A distributional reinforcement learning model for optimal glucose control after cardiac surgery,” npj Digital Medicine, vol. 8, Art. 313, 2025. doi:10.1038/s41746-025-01709-9.
34. U.S. Food and Drug Administration, “Artificial intelligence in software as a medical device,” FDA Digital Health Center of Excellence. [Online]. Available: https://www.fda.gov/medical-devices/digital-health-center-excellence/artificial-intelligence-software-medical-device

35. X. Wu, X. Zhao, W. Chen, Q. Chen, L. Kong, and P. Li, “A systematic review of continuous glucose monitoring sensors: Principles, core technologies and performance evaluation,” Sens. Actuators Rep., vol. 10, p. 100361, Dec. 2025.
36. M. Mansour, M. S. Darweesh, and A. Soltan, “Wearable devices for glucose monitoring: A review of state-of-the-art technologies and emerging trends,” Alex. Eng. J., vol. 89, pp. 224–243, Feb. 2024, doi: 10.1016/j.aej.2024.01.021.
37. Y. Liu, C. Liu, J. Zheng, C. Xu, and D. Wang, “Improving explainability and integrability of medical AI to promote health care professional acceptance and use: Mixed systematic review,” J. Med. Internet Res., vol. 27, p. e73374, Aug. 2025, doi: 10.2196/73374.
38. Z. Sadeghi, R. Alizadehsani, M. A. Cifci, S. Kausar, R. Rehman, P. Mahanta, P. K Bora, A. Almasri, R. S. Alkhawaldeh, S. Hussain, B. Alatas, A. Shoeibi, H. Moosaei, M. Hladík, S. Nahavandi, and P. M. Pardalos, “A review of explainable artificial intelligence in healthcare,” Comput. Electr. Eng., vol. 118, p. 109370, Aug. 2024, doi: 10.1016/j.compeleceng.2024.109370.
39. S. Pati, S. Kumar, A. Varma, B. Edwards, C. Lu, L. Qu, J. J. Wang, A. Lakshminarayanan, S. Wang, M. J. Sheller, K. Chang, P. Singh, D. L. Rubin, J. Kalpathy-Cramer, and S. Bakas, “Privacy preservation for federated learning in health care,” Patterns, vol. 5, no. 7, p. 100974, Jul. 2024, doi: 10.1016/j.patter.2024.100974.
40. R. Eden, I. Chukwudi, C. Bain, S. Barbieri, L. Callaway, S. de Jersey, Y. George, A.-D. Gorse, M. Lawley, P. Marendy, S. M. McPhail, A. Nguyen, M. Samadbeik, and C. Sullivan, “A scoping review of the governance of federated learning in healthcare,” npj Digit. Med., vol. 8, no. 1, p. 427, Jul. 2025, doi: 10.1038/s41746-025-01836-3.
41. V. Rahimzadeh, “US regulation of medical artificial intelligence and machine learning (AI/ML) research and development,” in Research Handbook on Health, AI and the Law, E. Elgar, ch. 16, 2023.
42. A. Mahajan, K. Heydari, and D. Powell, “Wearable AI to enhance patient safety and clinical decision-making,” npj Digit. Med., vol. 8, Art. 176, Mar. 2025, doi: 10.1038/s41746-025-01071-6.
43. A. Coronato, M. Naeem, G. De Pietro, and G. Paragliola, “Reinforcement learning for intelligent healthcare applications: A survey,” Artif. Intell. Med., vol. 109, p. 101964, Sep. 2020, doi: 10.1016/j.artmed.2020.101964.
44. B. Allen, “The promise of explainable AI in digital health for precision medicine: A systematic review,” J. Pers. Med., vol. 14, no. 3, p. 277, Mar. 2024, doi: 10.3390/jpm14030277.
45. M. S. Haleem, D. Katsarou, E. I. Georga, G. E. Dafoulas, A. Bargiota, L. Lopez-Perez, M. Rujas, G. Fico, L. Pecchia, D. Fotiadis, and the Gatekeeper Consortium, “A multimodal deep learning architecture for predicting interstitial glucose for effective type 2 diabetes management,” Sci. Rep., vol. 15, p. 27625, Jul. 2025, doi: 10.1038/s41598-025-07272-3.
46. A. Brauneck, L. Schmalhorst, M. M. K. Majdabadi, M. Bakhtiari, U. Völker, J. Baumbach, L. Baumbach, and G. Buchholtz, “Federated machine learning, privacy-enhancing technologies, and data protection laws in medical research: Scoping review,” J. Med. Internet Res., vol. 25, Mar. 2023, doi: 10.2196/41588.
47. R. Eden, K. S. Garg, G. D. Susskind, “Governance of federated learning in health and care: a scoping review,” J. Med. Internet Res., vol. 27, e57550, 2025. doi:10.2196/57550.
48. K. Bruynseels, F. Santoni de Sio, and J. van den Hoven, “Digital twins in health care: Ethical implications of an emerging engineering paradigm,” Frontiers in Genetics, vol. 9, Feb. 2018, doi: 10.3389/fgene.2018.00031.
49. R. E. Teixeira and S. Malin, “The next generation of artificial pancreas control algorithms,” J. Diabetes Sci. Technol., vol. 2, no. 1, pp. 105–112, Jan. 2008, doi: 10.1177/193229680800200115.
50. Q. F. Haoran Li Liang Gao, “An end-to-end decentralised scheduling framework based on deep reinforcement learning for dynamic distributed heterogeneous flowshop scheduling,” International Journal of Production Research, 2025, doi: 10.1080/00207543.2024.2449240.
51. I. V. Sergei V. Goncharov, “Reinforcement learning for urban public transport driver scheduling,” Modeling and Analysis of Information Systems, 2026, doi: 10.18255/1818-1015-2026-1-30-47.
52. C. S. T. Thangarasan M. Devika, “A deep reinforcement learning framework with explainable AI for personalized and interpretable treatment recommendations in healthcare,” CompSci & AI Advances, 2025, doi: 10.69626/cai.2025.0025.
53. C. K. M. Ziyang Lu M. C. Gursoy, “Explainable AI for radar resource management: Modified LIME in deep reinforcement learning,” 2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN), 2025, doi: 10.1109/ICMLCN64995.2025.11140538.
54. A. J. J. Beck, “Explainable AI (XAI) for arrhythmia detection from electrocardiograms,” arXiv.org, 2025, doi: 10.48550/arXiv.2508.17294.
55. M. D. B. S. Sonawane, “Risk-based validation of software, automation and artificial intelligence in pharmaceuticals,” Biosciences Biotechnology Research Asia, 2025, doi: 10.13005/bbra/3441.

56. P. F. Carvalho E Mascarenhas M, “Predetermined change control plans: Guiding principles for advancing safe, effective, and high-quality AI-ML technologies,” 2025, doi: 10.2196/76854.
57. B. R. Mehta V Komanduri A, “Evaluating transparency in AI/ML model characteristics for FDA-reviewed medical devices,” 2025, doi: 10.1038/s41746-025-02052-9.
58. B.-T. C. Amaya-Santos S Vargas C, “Frameworks encompassing intersectional perspective of artificial intelligence in healthcare. Scoping review,” 2025, doi: 10.1016/j.puhip.2025.100713.
59. G. S. Geukes Foppen RJ Gioia V, “Methodology for safe and secure AI in diabetes management,” 2024, doi: 10.1177/19322968241304434.