

REVIEW ARTICLE

Reinforcement and federated learning for privacy-preserving glucose regulation: A review

Sherif Abdelfattah*, Rahul Biswa Karma, Rohit Biswa Karma, and Babu Baniya

Department of Computer Science and Information Systems, Bradley University, Peoria, IL, United States of America

Abstract

Type 1 diabetes mellitus (T1DM) requires continuous monitoring of blood glucose levels and careful insulin dosing to maintain safe glycemic control. Reinforcement learning (RL) has emerged as a promising approach for automated glucose regulation in T1DM by enabling adaptive insulin dosing through sequential decision-making. However, real-world deployment remains constrained by safety requirements, inter-patient variability, limited clinical data, and strict privacy regulations that hinder centralized model development. This review aims to synthesize recent advances in RL-based automated insulin delivery (AID) and examine how federated learning (FL) can enable privacy-preserving personalization across distributed clinical and device data. We organize the literature into key themes, including: (i) online RL controllers developed primarily in simulation environments, (ii) offline RL methods that learn dosing policies from logged glucose-insulin trajectories (e.g., batch-constrained and conservative value-learning approaches), (iii) safety-aware strategies aimed at minimizing hypoglycemia risk and improving robustness under practical disturbances, and (iv) hybrid architectures integrating glucose prediction and control. We further discuss FL designs for personalization in heterogeneous, non-identically distributed patient data and highlight their relevance to diabetes management. Across the studies reviewed, most reinforcement learning approaches for glucose control are evaluated in simulation environments, while only a limited number use real-world clinical or continuous glucose monitoring (CGM) datasets. Federated learning studies are emerging for privacy-preserving personalization, and only a small subset of works currently explore hybrid RL-FL or federated reinforcement learning frameworks. Finally, we identify open challenges related to sim-to-real generalization, clinical validation, interpretability, scalability, and privacy risks in distributed learning, and outline future directions for robust, clinically viable next-generation diabetes management systems.

***Corresponding author:**
Sherif Abdelfattah
(sabdelfattah@bradley.edu)

Citation: Abdelfattah S, Karma RB, Karma RB, Baniya B. Reinforcement and federated learning for privacy-preserving glucose regulation: A review. *Artif Intell Health*. 2026;3(4):026050010. doi: 10.36922/AIH026050010

Received: January 31, 2026

Revised: March 9, 2026

Accepted: March 12, 2026

Published online: April 24, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Keywords: Reinforcement learning; Offline reinforcement learning; Federated learning; Privacy-preserving healthcare AI; Artificial pancreas; Diabetes management; Safe control

1. Introduction

Diabetes mellitus remains one of the most prevalent chronic diseases worldwide, and access to affordable treatment, including insulin, is essential for survival and long-term health outcomes.¹ Type 1 diabetes mellitus (T1DM), formerly known as juvenile diabetes

or insulin-dependent diabetes, is a chronic condition characterized by β -cell failure and insulin deficiency due to an autoimmune response.² As a result, regular insulin administration is required to maintain blood glucose levels within the normal range.³ For individuals living with T1DM, effective glycemic management requires continuous monitoring and frequent treatment decisions; when diagnosed with T1DM, an individual may make up to 180 decisions per day to regulate blood glucose levels.⁴ Insulin dosing must be carefully managed, as insufficient insulin can lead to hyperglycemia, resulting in long-term complications such as retinal damage⁵, cardiovascular disease⁶, and neuropathy.⁷ Conversely, excessive insulin administration can cause hypoglycemia, which may lead to loss of consciousness or death. Accurate glucose regulation depends on multiple factors, most notably blood glucose dynamics and caloric intake. These considerations generate at least 15–17 data points per day, which can increase substantially with continuous glucose monitoring (CGM) systems that produce up to 300 measurements per day, approximately one reading every five minutes.⁸

Another common form of diabetes is type 2 diabetes mellitus (T2DM), which is primarily caused by insulin

resistance and progressive pancreatic β -cell dysfunction, leading to impaired glucose uptake and dysregulated glycemic control.⁹ Lifestyle interventions and weight reduction may induce remission in some individuals, particularly in early-stage disease; however, T2DM remains a chronic condition and may recur or progress over time.¹⁰ Given the rising global burden of diabetes and its multi-organ complications, the development of scalable, accurate, and user-friendly strategies for glucose regulation remains a critical healthcare need.

Insulin administration can be broadly classified as manual injection or automated delivery via an automated insulin delivery (AID) system¹¹, which integrates a CGM device, an insulin pump, and a blood glucose control (BGC) algorithm that computes insulin dosing to regulate glucose levels in a closed-loop manner¹², as illustrated in Figure 1. Commercial AID systems have achieved substantial improvements in glycemic control, often reporting time-in-range (TIR) values of approximately 65–80% in evaluations.¹³ However, these systems still rely on manual meal announcements and user-provided exercise information to help prevent hypoglycemia, limiting autonomy and leaving room for further improvement.

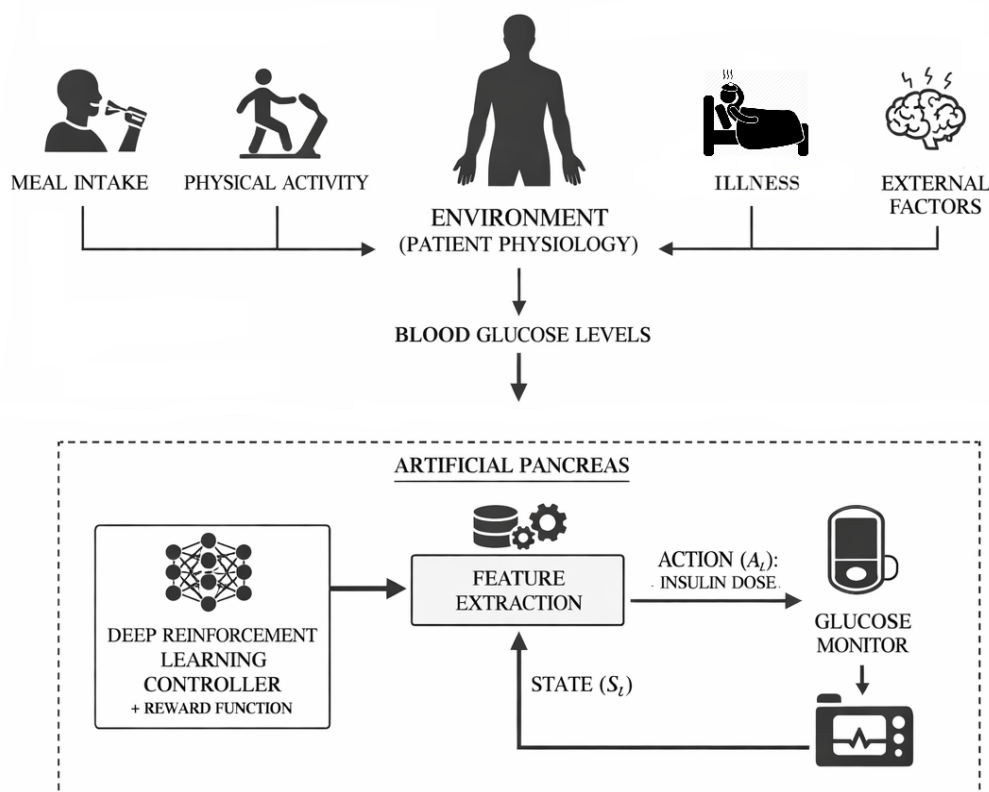


Figure 1. Reinforcement learning-based automated insulin delivery (AID) system architecture. Image created by the authors.

1.1. Terminology

Throughout this review, we use “automated insulin delivery (AID)” as the primary term to describe closed-loop systems that integrate a continuous glucose monitor, an insulin pump, and a control algorithm. The term “artificial pancreas (AP)” is widely used in the literature and is considered synonymous with AID in this context. “Hybrid closed-loop (HCL)” refers to systems that automate basal insulin delivery while still requiring user inputs, such as meal announcements, to maintain safety and performance. For consistency, we primarily adopt the term AID unless referencing terminology used in specific cited studies.

Within this context, reinforcement learning (RL) has emerged as a promising paradigm for automated glucose regulation and closed-loop insulin delivery. From a research perspective, automated insulin delivery has generated significant interest within the RL community¹¹, as RL formulates insulin dosing as a sequential decision-making problem in which an agent learns a policy that optimizes long-term glycemic outcomes. Unlike traditional controllers used in commercial systems, RL-based approaches can potentially adapt dosing strategies under uncertainty, accommodate inter-patient variability, and improve robustness to disturbances such as irregular meals, sensor noise, and imperfect user inputs. Recent work has explored simulation-trained online RL controllers, offline RL methods trained on logged glucose–insulin trajectories, and safety-aware strategies designed to reduce the risk of hypoglycemia.

In parallel, privacy-preserving learning approaches, such as federated learning (FL), have become increasingly relevant to healthcare AI, enabling collaborative model development across distributed devices and clinical sites without requiring the sharing of raw data. This is particularly important for diabetes management, where patient data are sensitive and heterogeneous and where privacy regulations can restrict centralized aggregation. Integrating FL into learning-based diabetes management pipelines may support scalable personalization while mitigating data-sharing risks; however, FL introduces challenges, including non-IID data distributions, client heterogeneity, communication constraints, and potential privacy leakage via shared model updates.

Although prior work has examined glucose control algorithms and machine learning methods for diabetes management, few studies systematically analyze reinforcement learning as a closed-loop control paradigm for AID systems, and even fewer address the emerging role of federated learning in privacy-preserving personalization. Moreover, existing surveys often treat RL and FL separately, with limited unified discussion of

safety constraints, evaluation protocols, and real-world deployment considerations. This underscores the need for an up-to-date synthesis of advances in learning-based control and privacy-preserving distributed learning frameworks in the context of diabetes care.

In this review, we provide a structured synthesis of reinforcement learning and federated learning approaches for diabetes management and automated insulin delivery. Specifically, we:

- (i) Summarize reinforcement learning formulations and algorithmic design choices used for insulin dosing and glucose regulation
- (ii) Review learning-based control strategies, including online and offline reinforcement learning approaches as well as safety-aware designs aimed at reducing hypoglycemia risk.
- (iii) Examine federated learning architectures and personalization strategies relevant to diabetes applications under heterogeneous patient data.
- (iv) Discuss how ideas from reinforcement learning and federated learning research may translate to distributed and privacy-preserving diabetes management systems.
- (v) Outline key challenges related to evaluation settings, safety considerations, and the practical deployment of learning-based glucose control systems.

We conducted a semi-systematic literature review to examine how RL and FL are currently used, or could reasonably be used, for diabetes management and closed-loop blood glucose control. Searches were conducted in IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, PubMed, arXiv, and related sources, supplemented by backward reference checking to identify additional key studies. We focused primarily on peer-reviewed journal and conference publications from 2010 to 2025, using combinations of search terms such as reinforcement learning, offline reinforcement learning, federated learning, AID, blood glucose control, closed-loop diabetes, insulin dosing, and privacy-preserving healthcare. Study selection followed a two-stage screening process (titles/abstracts followed by full-text review), with inclusion determined by each paper’s direct relevance to RL/FL-based glucose regulation, transferable control-learning design, or privacy-aware system development; purely supervised learning studies without decision-making components and unrelated healthcare topics were excluded. The included literature was then categorized by contribution type, covering online/offline/modular/multi-agent RL formulations; FL architectures and aggregation strategies; privacy/security mechanisms; and hybrid RL–FL frameworks, with the extraction of evaluation settings and clinically meaningful metrics (e.g., time-in-range

and hypoglycemia-related indicators) where available. After screening and deduplication, a total of 114 studies were included in this review, spanning reinforcement learning-based glucose control, FL and privacy-preserving healthcare analytics, and hybrid federated reinforcement learning frameworks, with the majority of publications appearing after 2019. All included studies are listed in the References section, and the main text highlights representative studies across categories.

Broadly, existing work on learning-based diabetes management can be grouped into four recurring approaches. The first focuses on online reinforcement learning controllers trained primarily in simulation environments for closed-loop insulin delivery. The second explores offline RL methods that learn dosing policies from previously collected glucose–insulin trajectories in order to reduce unsafe exploration. A third line of work emphasizes safety-aware and modular architectures, in which learning-based policies are combined with rule-based safeguards, prediction components, or risk-sensitive objectives. Finally, distributed and privacy-preserving frameworks based on FL have emerged to support collaborative training and personalization under heterogeneous patient data. Throughout this review, we synthesize findings across these directions to highlight design choices that consistently improve glycemic outcomes while maintaining safety and deployment feasibility.

The remainder of this paper is organized as follows. Section 2 introduces the fundamentals of RL relevant to glucose regulation. Section 3 reviews reinforcement control algorithms for closed-loop insulin delivery and discusses their design considerations and safety requirements. Subsequent sections synthesize online and offline RL approaches, hybrid prediction-control frameworks, and evaluation protocols for AID systems. We then examine FL methods for privacy-preserving personalization and discuss emerging directions in federated reinforcement learning in Section 4. Finally, we summarize key findings, discuss limitations, and outline future research directions.

2. Reinforcement learning

2.1. Background

Reinforcement learning (RL) is a machine learning approach in which an agent interacts with an environment and learns decision-making strategies through feedback in the form of rewards.¹⁴ The primary objective of an RL algorithm is to learn behaviors that achieve specific goals by maximizing the cumulative reward over time. The Markov Decision Process framework is commonly used

to formulate the reinforcement learning problem, which is defined by the tuple (S, A, P, R, γ) . In this framework, S denotes the state space, A denotes the action space, P denotes the state transition probabilities, R denotes the reward function, and γ is the discount factor, whose value lies in $[0,1]$, determining the relative importance of immediate versus future rewards.¹⁵

Within this framework, reinforcement learning aims to learn a policy that maps observed states to actions to maximize cumulative reward.¹⁶ The optimal policy can be defined as the policy that maximizes the expected cumulative discounted reward over time and is commonly expressed as:

$$\pi^* = \arg \max_{\pi \in \Pi} E \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(s_t)) \right] \quad (1)$$

Here, Π denotes the set of all possible policies, and s_t represents the state of the environment at time step t . The policy π maps observed states to actions, and the objective in **Equation 1** is to identify an optimal policy π^* that maximizes the expected cumulative discounted reward over time.

The reward function $R(s_t, \pi(s_t))$ evaluates the immediate outcome of executing the action selected by the policy in state s_t , while the discount factor $\gamma \in [0,1]$ determines the relative importance of future rewards.

This formulation represents a standard Markov Decision Process objective and underpins many RL algorithms. **Figure 2** illustrates this Markov Decision Process formulation for an RL-based AID, highlighting the interaction between patient state observations, insulin dosing actions, reward feedback, and the underlying glucose–insulin dynamics.

Across these formulations, the key challenge for glucose regulation is not policy optimality alone, but achieving stable and safe control under delayed insulin dynamics, physiological uncertainty, and strict hypoglycemia constraints. Unlike standard RL benchmarks, AID systems must balance long-horizon performance with real-time safety, robustness, and patient-specific variability. These domain-specific requirements fundamentally shape state design, reward engineering, and the incorporation of safety mechanisms in reinforcement learning-based insulin control.

In practice, this suggests that reinforcement learning methods for glucose regulation should be evaluated not only by cumulative reward, but also by how effectively they handle delayed insulin effects, enforce safety constraints, and remain robust to variability and distribution shifts.

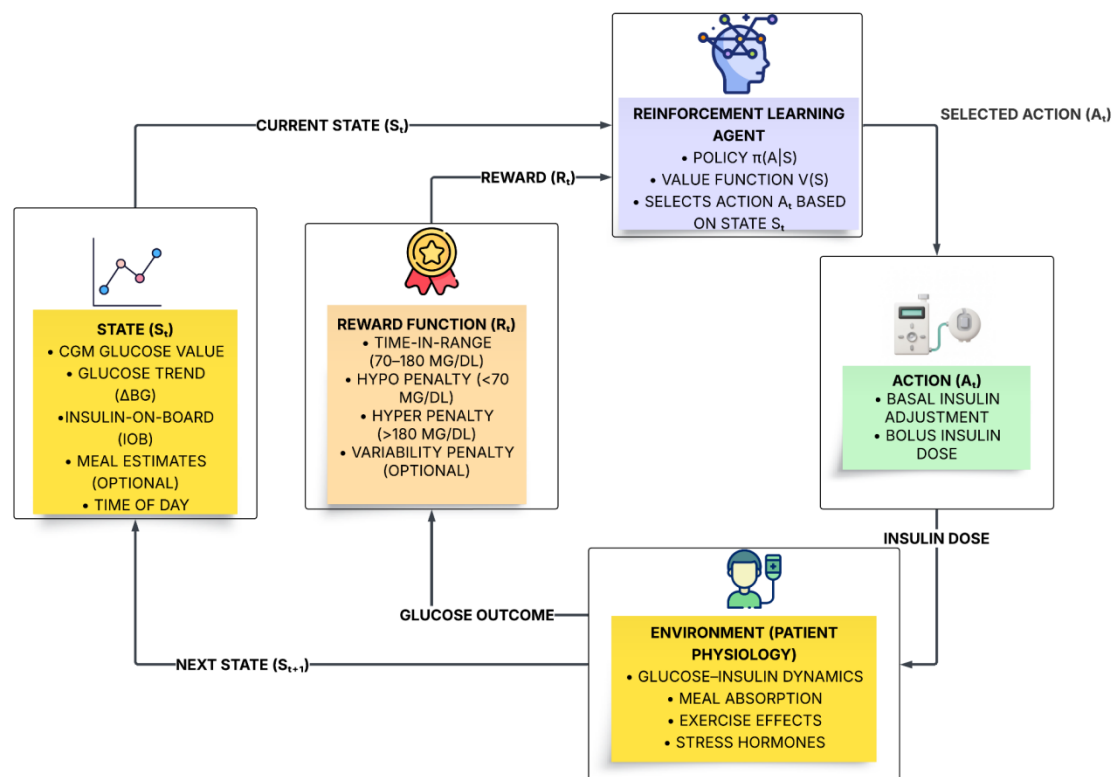


Figure 2. Markov Decision Process representation of the reinforcement learning-based automated insulin delivery (AID). Image created by the authors.

2.2. Models and simulators

The next step in the development of an AID system involves the deployment of mathematical modeling and computer-based simulation. A well-designed simulator can provide valuable insights into the safety and limitations of closed-loop glucose control systems while enabling cost-effective exclusion of ineffective control scenarios.¹⁷ In addition to safety assessment, such simulation environments must capture real-world glucose dynamics, sensor imperfections, and inter-patient variability in T1DM to provide a rich and realistic environment for robust controller training.

In addition to invasive sensors and simulators, recent work has explored non-invasive glucose estimation using multimodal wearable signals and interpretable machine learning. Shan and Yu¹⁸ proposed a framework that combines photoplethysmography, electrodermal activity, skin temperature, and food log features to predict blood glucose levels without direct blood sampling. In a dataset of 16 adults monitored over multiple days, the model trained with LightGBM achieved a root mean square error (RMSE) of 12.9 mg/dL, a mean absolute relative difference (MARD) of 7.9%, and an mean absolute error (MAE) of 8.82 mg/dL, with feature relevance analysis showing that

recent carbohydrate intake and short-term physiological changes were among the most important predictors. While this study focuses on glucose prediction rather than control, integrating such non-invasive multimodal features into state representations could improve RL-based controllers and reduce reliance on invasive devices, especially when interpretability and patient comfort are important.

Over the past several years, a variety of models have been proposed to replicate glucose–insulin dynamics under different physiological assumptions. In the context of RL, glucose regulation is commonly formulated as a Markov Decision Process, where the controller interacts with the environment through states, actions, transitions, and rewards.^{19,20} At each decision step, the agent observes the patient's physiological state and selects an insulin-related action, such as a basal rate adjustment or a bolus recommendation.¹⁷ The environment then evolves according to the underlying glucose–insulin dynamics, producing a new physiological state and a delayed reward that reflects metabolic safety or glucose-tracking performance.

The actual transition dynamics governing human glucose metabolism are not directly observable; therefore,

agents must learn these dynamics through interaction with either simulated or real environments.¹⁷ This sequential decision-making process highlights how insulin dosing decisions are continuously adapted based on glucose measurements and prior insulin actions. In continuous-control settings, both the state and action spaces are represented as real-valued vectors, and simulators propagate system dynamics forward in discrete time. Reward functions are typically bounded or sparse and are designed to promote smooth glucose regulation while penalizing hypoglycemia and hyperglycemia.¹⁹ A wide range of physiological and data-driven models have been developed for RL-based diabetes management. The models differ in their mathematical structure, underlying physiological assumptions, computational complexity, and degree of personalization.²¹

One of the most widely used approaches is the model proposed by Ferdinando *et al.*,²² which employs delay differential equations to capture temporal delays in subcutaneous insulin absorption. Insulin kinetics in this model are represented by first-order dynamics, whereas glucose dynamics evolve in response to insulin action. This enables estimation of insulin-on-board and supports dosing decisions that depend on both the timing and magnitude of insulin delivery. By varying model parameters, diverse clinical scenarios can be generated to reflect patient heterogeneity.²²

At the opposite end of the model-complexity spectrum lies the Lehmann–Deutsch physiological model²³, which employs a detailed multi-compartment representation of T1DM physiology. This model uses plasma glucose concentration, insulin absorption kinetics, hepatic glucose balance, renal glucose excretion, insulin distribution volume, and exogenous insulin administration. The patients are characterized by distinct parameter sets, enabling the model to capture inter-patient variability and reduce errors arising from noise or overly simplified state representations. While this level of detail enables testing under different conditions, careful calibration of parameters is required to ensure long-term simulation stability.

In contrast to mechanistic models, Shifrin’s transition probability-based model²⁴ treats patient physiology as a stochastic process governed by a transition matrix. The model encodes the probability of transitioning between behavioral and physiological states, including glucose levels, carbohydrate intake, and time since the last insulin action. This probabilistic formulation enables the integration of variability arising from uncontrollable factors such as hormonal disturbances, unannounced meals, stress, and behavioral changes. By emphasizing stochasticity rather

than strict physiological structure, this model is well-suited for training agents that must operate under uncertainty or partial observability, where deterministic models may fail to capture real-world irregularities. The resulting agents demonstrate improved robustness and safety across unpredictable conditions.

Another modeling approach is Lee *et al.*’s pharmacokinetic and pharmacodynamic (PK/PD) model²⁵, which emphasizes insulin time–action profiles described by gamma distributions. In this formulation, the distribution parameters control the shape and time-to-peak insulin activity. Model calibration is informed by clamp-test data, in which insulin is infused subcutaneously under controlled conditions, and a proportional–integral controller adjusts intravenous glucose infusion to maintain euglycemia.²⁶ Temporal discount factors are derived through normalization procedures, enabling the model to account for delayed and long-lasting insulin effects, as well as the downstream consequences of insulin decisions.²⁵ This approach allows for a high degree of personalization based on patient-specific insulin kinetics.

The Hovorka glucoregulatory model²⁷ is another widely used physiological simulator. This model preserves statistical correlations derived from healthy subject data and samples physiological parameters from log-normal distributions to reflect natural variability. Parameters oscillate at random frequencies and phases, enabling simulation of both slow and fast meal dynamics via modulated carbohydrate absorption. As a result, the model captures everyday variability, noise, and dynamic disturbances that closely resemble real clinical scenarios.²⁷

In 2008, computer simulation models were formally accepted as a substitute for animal testing in the preclinical evaluation of closed-loop insulin delivery systems.¹⁷ This work led to the development of the UVA/Padova metabolic simulation framework, which describes interactions among plasma glucose, tissue glucose, and plasma insulin through key metabolic fluxes, including endogenous glucose production, glucose appearance from meals, glucose utilization, renal extraction, and insulin absorption and degradation.²¹

Within this framework, glucose kinetics are commonly modeled using a two-compartment model comprising a plasma compartment and a remote tissue compartment. Insulin-independent glucose utilization occurs primarily in the plasma compartment, while insulin-dependent utilization occurs in remote tissues. The glucose dynamics are expressed as:

$$\dot{G}_p(t) = EGP(t) - U(t) - k_1 G_p(t) + k_2 G_t(t) \quad (2)$$

$$\dot{G}_t(t) = -k_2 G_t(t) + k_1 G_p(t) \quad (3)$$

In **Equations 2** and **3**, $G_p(t)$ and $G_t(t)$ denote plasma and tissue glucose concentrations, respectively. The term $EGP(t)$ represents endogenous glucose production, while $U(t)$ denotes total glucose utilization. The parameters k_1 and k_2 describe glucose exchange rates between the plasma and tissue compartments. The dot notation ($\dot{\cdot}$) indicates time derivatives.¹⁷

Insulin-independent glucose utilization U_{ii} is treated as a constant, while insulin-dependent glucose utilization U_{id} is modeled using a Michaelis–Menten relationship, as defined in **Equation 4**.²⁷

$$U_{id}(t) = \left(\frac{V_{m0} + V_{mx} X(t)}{K_{m0} + G_t(t)} \right) \cdot G_t(t) \quad (4)$$

Here, V_{m0} represents the basal glucose uptake capacity, V_{mx} denotes insulin sensitivity, $X(t)$ is the insulin action variable, and K_{m0} is the Michaelis constant controlling saturation behavior.

Insulin action in the tissue compartment evolves according to **Equation 5**.²⁷

$$\dot{X}(t) = -p_{2U} X(t) + p_U (I(t) - I_b) \quad (5)$$

In **Equation 5**, $X(t)$ represents insulin action, $I(t)$ denotes the plasma insulin concentration, I_b is the basal insulin level, and the parameters p_U and p_{2U} control insulin activation and decay rates, respectively.²⁷

Endogenous glucose production is modeled as:

$$EGP(t) = k_{p1} - k_{p2} G_p(t) - k_{p3} X_L(t) \quad (6)$$

$$\dot{X}_L(t) = -k_i X_L(t) + k_L (I(t) - I_b) \quad (7)$$

Here, $X_L(t)$ denotes delayed insulin action on glucose production, while the parameters k_{p1} , k_{p2} , k_{p3} , k_i , and k_L govern suppression of glucose production and insulin delay effects.²⁷

The simulator includes a virtual population of 300 subjects, consisting of 100 adults, 100 adolescents, and 100 children. Each subject is generated using a joint physiological parameter distribution, with inter-subject variability preserved through a shared covariance matrix. Group-specific adjustments, such as increased insulin sensitivity in children and reduced sensitivity in adolescents, are applied. Subsequent extensions of the model incorporated improved gastric emptying dynamics, refined glucose appearance models, and expanded clinical validation.^{17,21}

Overall, RL provides a principled framework for

modeling insulin dosing as a sequential decision-making process and for learning adaptive control policies that optimize long-horizon glycemic outcomes under uncertainty. Because direct exploration in real patients is infeasible and unsafe, most RL-based AID controllers rely on physiologically grounded mathematical models and *in silico* simulators to support training, benchmarking, and safety evaluation prior to clinical translation. These foundations motivate the reinforcement control algorithms reviewed in the next section, which operationalize RL principles through practical policy optimization strategies and safety-aware learning designs tailored to closed-loop glucose regulation.

3. Reinforcement control algorithm

Across the modeling and simulation literature, several practical design principles repeatedly appear in successful RL-based controllers. Most approaches incorporate insulin history or insulin-on-board estimates to address delayed physiological effects, employ clinically motivated reward functions that penalize hypoglycemia and extreme glucose excursions, and impose explicit action limits or safety layers to prevent unsafe dosing. In addition, robust evaluation under meal variability, sensor noise, and parameter uncertainty is commonly used to assess real-world feasibility. Together, these recurring elements help explain performance differences across algorithms and form a foundation for effective closed-loop learning systems.

Reinforcement learning control algorithms provide practical mechanisms for implementing learned insulin dosing policies in automated insulin delivery systems. In contrast to classical controllers, which are typically tuned offline, RL-based controllers adapt decision-making policies using interaction data while explicitly optimizing long-horizon objectives. However, their clinical use requires integration with safety mechanisms that constrain insulin delivery and reduce the risk of hypoglycemia. **Figure 3** presents a representative safety-aware control architecture and provides system-level context for the RL control algorithms reviewed in this section, including policy-gradient methods such as Proximal Policy Optimization (PPO), offline RL approaches trained from historical trajectories, and value-based methods from the Deep Q-Networks (DQN) family, along with their associated safety and deployment considerations.

Across the literature, improvements in glucose regulation are rarely achieved through RL alone, but rather through the integration of learning algorithms with safety mechanisms, carefully designed reward structures, and state representations that account for delayed insulin

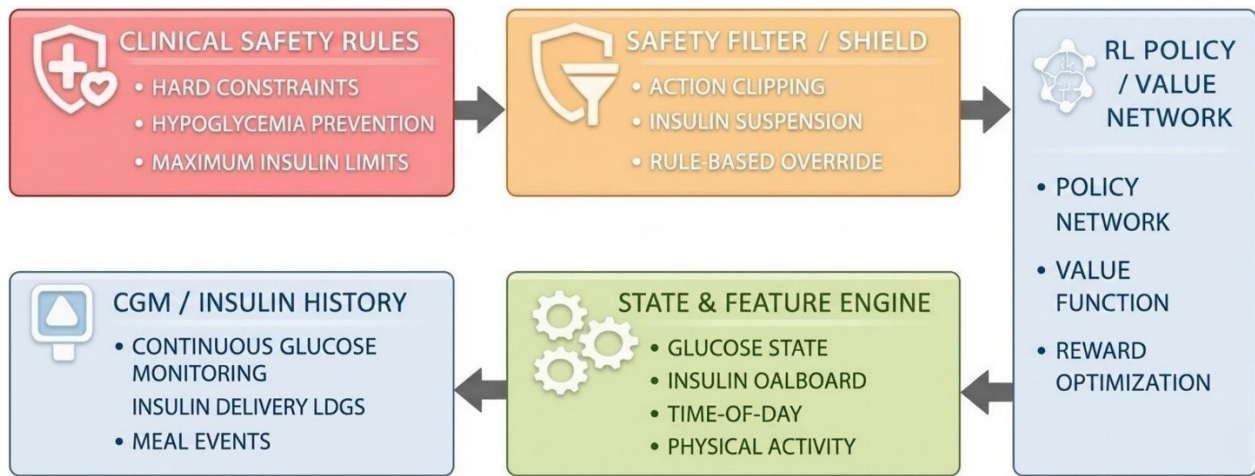


Figure 3. Safety-aware reinforcement learning control architecture for automated insulin delivery. Image created by the authors. Abbreviation: CGM: Continuous glucose monitoring.

dynamics.

3.1. Proximal Policy Optimization (PPO)

Proximal Policy Optimization (PPO) is a policy-gradient RL algorithm widely used in continuous control tasks due to its stable training behavior and relatively simple implementation. The method is particularly attractive for insulin delivery because it supports smooth policy improvement while avoiding extensive updates that could lead to unstable or unsafe action changes. PPO operates in an iterative loop: data are collected by interacting with the environment, and policy parameters are then updated using gradient ascent on a surrogate objective. To improve stability, PPO constrains updates so that the new policy remains close to the previous policy.²⁸

The fundamental policy-gradient objective is defined as:

$$L_{PG}(\theta) = \hat{E}_t \left[\log \log \pi_\theta(a_t | s_t) \cdot \hat{A}_t \right] \quad (8)$$

In **Equation 8**, $\pi_\theta(a_t | s_t)$ denotes the stochastic policy parameterized by θ , which defines the probability of selecting action a_t in state s_t . The advantage estimate $\hat{A}_t$ reflects the extent to which the selected action differs from the expected behavior at that state.

The expectation $\hat{E}_t$ is computed over samples collected from the environment and is typically approximated using a finite batch of trajectories. PPO optimizes a clipped surrogate objective function $L_{CLIP}(\theta)$ given by:

$$L_{CLIP}(\theta) = \hat{E}_t \left[\min \left(r_t(\theta) \cdot \hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) \cdot \hat{A}_t \right) \right] \quad (9)$$

The probability ratio $r_t(\theta)$ is defined as:

$$r_t(\theta) = \frac{\pi_\theta(a_t | s_t)}{\pi_{\theta_{old}}(a_t | s_t)} \quad (10)$$

This ratio compares the likelihood of selecting an action a_t under the updated policy with that under the previous policy. The clipping operation confines $r_t(\theta)$ to the interval $[1-\epsilon, 1+\epsilon]$, where ϵ is a small positive constant typically chosen between 0.1 and 0.3. By doing so, PPO discourages abrupt policy changes while still allowing consistent performance improvements over time.

The complete PPO objective combines the clipped policy loss with auxiliary terms that support stable value learning and sufficient exploration:

$$L_{PPO}(\theta) = E_t \left[L_{CLIP}(\theta) - c_1 (V_\theta(s_t) - V_t^{target})^2 + c_2 S[\pi_\theta](s_t) \right] \quad (11)$$

In **Equation 11** $V_\theta(s_t)$ denotes the estimated value of the current state s_t under the policy being learned, while V_t^{target} represents the target value obtained from Monte Carlo returns or generalized advantage estimation. The squared difference between these terms penalizes value prediction errors and improves training stability. The entropy term $S[\pi_\theta](s_t)$ encourages exploration by preventing premature convergence to deterministic policies, particularly during early training. Owing to its balance between learning efficiency and robustness, PPO is commonly used as a baseline in RL studies of both discrete and continuous control problems, including automated insulin delivery.

3.1.1. Dual Policy Optimization (Dual PPO)

The work presented by Marchetti *et al.*²⁹ is motivated by the need to increase automation in insulin delivery systems for individuals with T1DM. Existing glucose management

strategies rely heavily on patient input, thereby limiting the extent to which insulin delivery can operate autonomously. To reduce this dependency, the study proposes a Dual Policy Optimization (Dual PPO) approach implemented within a hybrid closed-loop insulin delivery framework.

The proposed control strategy comprises two independently trained PPO agents operating under different insulin-delivery constraints. A high-capacity agent (CapH) and a low-capacity agent (CapL), with $\text{CapL} < \text{CapH}$, are each assigned an upper bound on insulin infusion rates. System operation is divided into three glucose-dependent regions. When glucose levels exceed a predefined threshold, the high-capacity agent is activated, enabling more aggressive insulin delivery. As glucose levels approach the safe range, control is transferred to the low-capacity agent, which applies more conservative dosing. When glucose levels approach hypoglycemic risk, a safety region is entered in which insulin delivery is suspended entirely.

This region-based, dual-agent design provides a structured mechanism for adapting control aggressiveness across glycemic regimes while enforcing explicit safety constraints. By combining capacity-limited policies with RL-based control, the Dual PPO framework enhances automation and robustness while maintaining safeguards against hypoglycemia.

3.1.2. Safe-enhanced PPO for fully closed-loop AID

A PPO-based controller has also been proposed for fully closed-loop glucose regulation without requiring prior meal announcements. The method aims to improve

stability, safety, and sample efficiency through training refinements, including bounded action spaces to constrain insulin delivery within clinical limits, observation and advantage normalization, entropy annealing, learning-rate scheduling, reward normalization, gradient clipping, and orthogonal network initialization. In addition, a failure-aware advantage estimator is employed to reduce the influence of unsafe trajectories during training.

Safety is incorporated through a two-level mechanism. A proactive component shapes the reward using a clinically motivated risk metric, such as the Clarke error grid, penalizing glucose values outside the 70–180 mg/dL range. A reactive component applies rule-based constraints that scale or suspend insulin delivery based on current glucose levels, observed trends, and time-of-day considerations. For example, insulin delivery may be suspended at approximately 80 mg/dL when a downward trend is detected and then gradually reintroduced.

Training and evaluation are conducted using the Simglucose²¹ simulator with adult, adolescent, and pediatric populations under unannounced and randomized meal scenarios, including sensor artifacts. Performance is compared against proportional–integral–derivative (PID) controllers and other deep reinforcement learning baselines under the same evaluation protocols. For deployment, the design includes per-patient normalization procedures and transparent safety rules. Remaining challenges identified in the study include model interpretability, generalization across patients, and computational efficiency on embedded hardware.³⁰ [Table 1](#) summarizes the key differences among standard PPO, Dual

Table 1. Comparison of PPO-based control strategies for automated insulin delivery systems

Aspect	Standard PPO ²⁷	Dual PPO ²⁸	Safe-enhanced PPO ²⁹
Level of automation	Partial/baseline	Hybrid closed-loop	Fully closed-loop
Insulin constraints	Learned implicitly	Explicit capacity limits (CapL, CapH)	Explicit clinical bounds
Safety handling	Reward-based only	Region-based policy switching	Reward shaping + rule-based safety
Meal announcement required	Often required	Sometimes required	Not required
Hypoglycemia prevention	Indirect	Conservative low-capacity mode	Explicit insulin suspension and scaling
Interpretability	Low	Medium (capacity regions)	Higher (transparent safety rules)
Deployment readiness	Experimental	Intermediate	Higher for clinical settings
Evaluation cohort (n)	Simulation	Simulation cohort (virtual patients)	Simglucose: adult, adolescent, pediatric (virtual cohort)
Evaluation setting	<i>In silico</i>	<i>In silico</i>	<i>In silico</i>
Evaluation duration	Multi-day simulation episodes	Multi-day simulation	Multi-day randomized meal scenarios
Tested in real patients?	No	No	No

Abbreviations: CapH: High-capacity agent; CapL: Low-capacity agent; Dual PPO: Dual Policy Optimization; PPO: Proximal Policy Optimization.

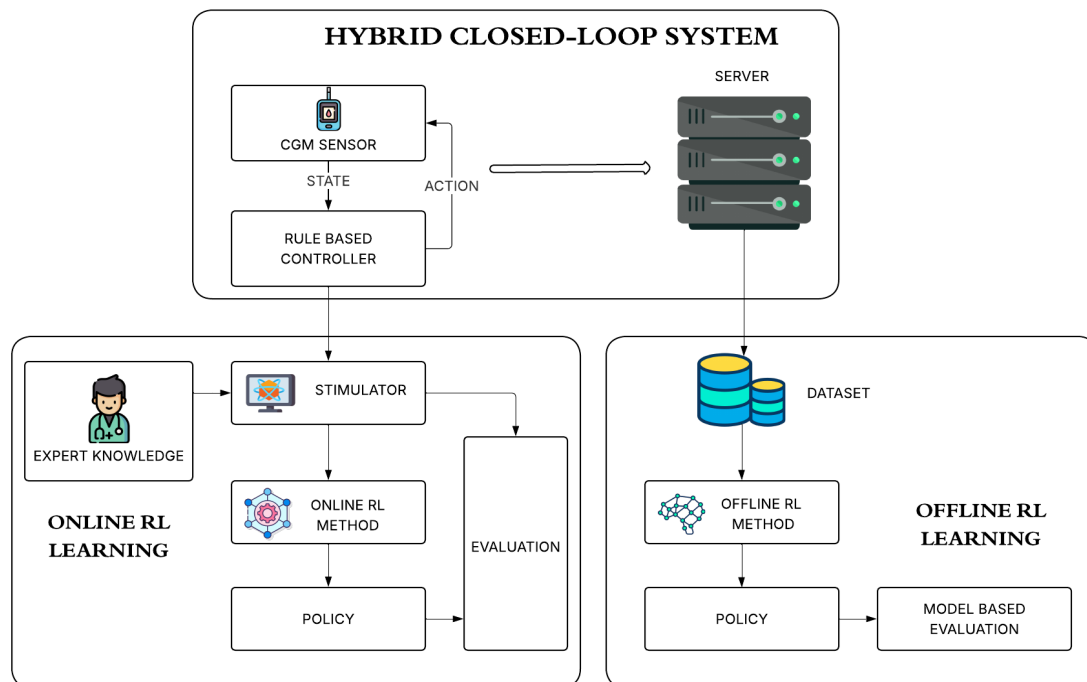


Figure 4. Overall architecture of the hybrid closed-loop reinforcement learning framework. Image created by the authors.
Abbreviations: CGM: Continuous glucose monitoring; RL: Reinforcement learning.

PPO, and safe-enhanced PPO controllers with respect to automation level, insulin constraints, safety handling, and reliance on meal announcements.

3.2. Offline reinforcement learning

In RL, an agent aims to learn a policy that maps environment states to actions in order to maximize cumulative reward. This is commonly achieved using a learned action–value (Q) function. In the standard RL setting, the agent interacts with the environment, continuously collects experience using its current policy, and updates the policy based on newly observed transitions.³¹ This type of online interaction can be impractical in many real-world applications due to the cost, time requirements, and potential safety risks associated with data collection.³²

Figure 4 illustrates a hybrid closed-loop glucose control framework that integrates both online and offline reinforcement learning pipelines. The upper module represents real-time closed-loop operation using CGM sensors and rule-based controllers, while the lower modules depict online RL training using a simulator and offline RL training based on historical datasets.

To overcome the limitations of online interaction, offline RL has been introduced. Over the past decades, significant progress in machine learning has been driven by data-driven approaches that rely on previously collected

datasets.³¹ A similar concept applies to offline RL, where policy learning is performed using only logged data and does not require interaction with the environment during training.³³ In this setting, the agent learns entirely from an offline dataset, which improves safety by avoiding risky exploration. However, challenges remain when the agent encounters out-of-distribution (OOD) states, as this can lead to overly optimistic Q-value estimates and potentially unsafe decisions.³⁴

While Figure 4 presents a system-level hybrid closed-loop control architecture, Figure 5 compares online, offline, and hybrid reinforcement learning paradigms with respect to data sources, training workflows, and deployment strategies. Figure 5 illustrates this hybrid paradigm, combining offline policy learning from historical glucose–insulin data with simulation-based online refinement and safety-aware closed-loop deployment.

3.2.1. Offline RL for hybrid closed-loop glucose control (BCQ, CQL, TD3-BC)

Emerson *et al.*³³ evaluates offline reinforcement learning approaches using the UVA/Padova simulator (see Section 2.2) with a cohort of 30 virtual patients. Model-free offline RL techniques such as BCQ (Batch-Constrained Deep Q-learning), CQL (Conservative Q-learning), and TD3-BC (Twin Delayed DDPG with Behavioral Cloning)

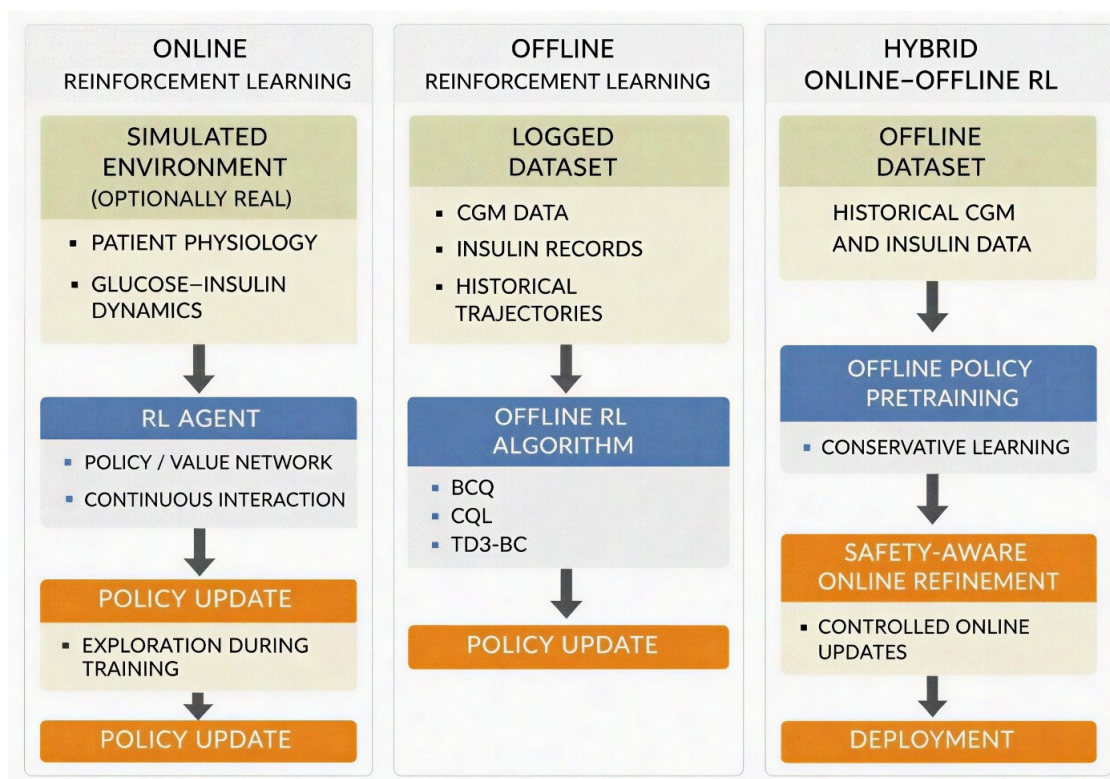


Figure 5. Comparison of online, offline, and hybrid reinforcement learning paradigms for closed-loop insulin delivery and glucose regulation. Image created by the authors.

Abbreviations: BCQ: Batch-Constrained Deep Q-learning; CQL; Conservative Q-learning; CGM: Continuous glucose monitoring; RL: Reinforcement learning; TD3-BC: Twin Delayed DDPG with Behavioral Cloning.

are employed to learn basal insulin dosing policies solely from logged trajectories. By remaining close to the behavior distribution (BCQ, TD3-BC) or enforcing conservative value estimates (CQL), these methods reduce off-support overestimation. Among them, TD3-BC demonstrated notable improvements in time-in-range compared to PID control, without increasing time-below-range, while requiring substantially less data than typical online RL methods. The approaches also demonstrated robustness to practical challenges, including missing CGM data, bolus misestimation, irregular meals, and compression artifacts.³⁵

For clarity, the core mechanisms underlying each method can be summarized as follows:

- **BCQ:** Constrains the learned policy to the dataset supported by sampling variational autoencoder (VAE)-generated near-dataset actions with small perturbations, using clipped double Q-learning to limit overestimation.³⁴
- **CQL:** Learns a conservative critic by penalizing Q-values for out-of-distribution state-action pairs

while increasing values for in-dataset samples, resulting in lower-bound value estimates that favor in-distribution actions.³⁵

- **TD3-BC:** Incorporates a behavior cloning term into the TD3 actor loss to keep actions close to the dataset, while twin critics and target policy smoothing mitigate overestimation.³⁶

3.2.2. PAINT: Preference adaptation for insulin control in T1DM

PAINT is an offline RL approach that learns flexible insulin-dosing policies from human preferences while explicitly enforcing safety constraints.³⁷ In this framework, patients or simulated users provide preference-based reward sketches derived from past glucose and insulin trajectories. A reward model is trained on these sketches and subsequently used to relabel the offline dataset. Policy optimization is then performed by fine-tuning a TD3 + BC controller under an explicit safety prior.

In standard TD3 + BC, the actor loss includes a behavior-cloning regularization term of the form:

$$(\pi(s) - a)^2 \quad (12)$$

where $\pi(s)$ denotes the action proposed by the learned policy at state s , and a represents the action observed in the logged dataset. This term encourages the learned policy to remain close to the behavior policy that generated the data.

In PAINT, this behavior-cloning term is replaced with a safety-aware regularization term:

$$(\pi(s) - \pi_{prior}(s))^2 \quad (13)$$

In Equation 13, $\pi_{prior}(s)$ denotes a predefined safety prior policy that encodes conservative insulin-dosing behavior.

By penalizing deviations from the safety prior rather than from the dataset actions, the learned policy is guided toward safe behavior while still allowing adaptation based on human preferences. A scalar parameter λ controls the relative influence of the preference-based reward signal during optimization, balancing preference satisfaction against safety preservation.

An experimental evaluation is conducted using the UVA/Padova simulator, with a PID controller serving as the baseline, and the Magni risk metric is used for safety assessment. To stabilize reward model training under imbalanced or discrete preference labels, stratified sampling across reward bins is employed. The results demonstrate reduced risk, improved post-meal time-in-range, and robustness to noisy or limited preference data.³⁷

Taken together, offline RL studies for glucose control reveal a consistent balance between safety and flexibility. Methods that remain close to observed behavior, such as BCQ and TD3-BC, tend to provide more stable and conservative insulin dosing, while approaches that enforce conservative value estimates reduce over-optimism but can become overly cautious when data coverage is sparse. A recurring limitation is that most evaluations still rely heavily on simulator-based rollouts. Future progress will likely depend on improved off-policy evaluation techniques and broader validation using real-world CGM and insulin pump datasets.

Taken together, offline RL approaches for glucose control show steady progress toward safer insulin-dosing strategies, but comparisons across studies should be made carefully. For example, Emerson *et al.*³³ evaluated BCQ, CQL, and TD3-BC using the UVA/Padova simulator (see Section 2.2) with a 30-patient virtual cohort. They report improvements in time-in-range compared to PID control, without increasing time-below-range, and show robustness

under simulated disturbances.

PAINT, on the other hand, builds on TD3 + BC by adding a safety-prior regularization term and evaluates performance using risk-sensitive measures such as the Magni index under preference-based reward relabeling. Although both studies use the same simulator environment, they differ in reward design, safety constraints, and dataset construction. These differences can directly affect reported TIR and risk outcomes.

In addition, most offline RL evaluations still rely on simulator-generated data rather than independent real-world cohorts. As a result, improvements in TIR or safety metrics should be interpreted within the context of the specific simulator setup and evaluation protocol, rather than treated as directly comparable benchmarks. More consistent reporting of datasets, reward definitions, and validation settings would improve cross-study comparison in this area.

3.3. Value-based reinforcement learning (Q-learning/DQN family)

Value-based reinforcement learning methods aim to learn an action-value function $Q(s, a)$ and select actions by acting greedily with respect to the estimated values. Deep Q-Networks (DQN) extend this framework by using a neural network as a function approximator and are commonly stabilized with an experience replay buffer and a separate target network.³⁸ Several extensions have been proposed to address limitations of the original DQN formulation. Double DQN mitigates value overestimation bias³⁹, dueling architectures decompose the Q-function into state-value and action-advantage components⁴⁰, prioritized experience replay increases the sampling frequency of transitions with high temporal-difference error⁴¹, multi-step return formulations improve credit assignment over longer horizons⁴², and distributional Q-learning models the full return distribution rather than only its expected value.⁴³

Figure 6 presents a DQN-based closed-loop glucose control framework that illustrates how state information, action decisions, reward feedback, and the environment interact. In glucose regulation, the system evolves slowly, and insulin's effects are delayed, which makes state design particularly important. To address this, the state is commonly extended with information about past insulin delivery. This is typically done using exponentially decayed summaries of recent actions or explicit insulin-on-board (IOB) estimates. Incorporating such history helps the system better approximate the Markov property and improves the reliability of value-based control approaches for diabetes management.⁴⁴

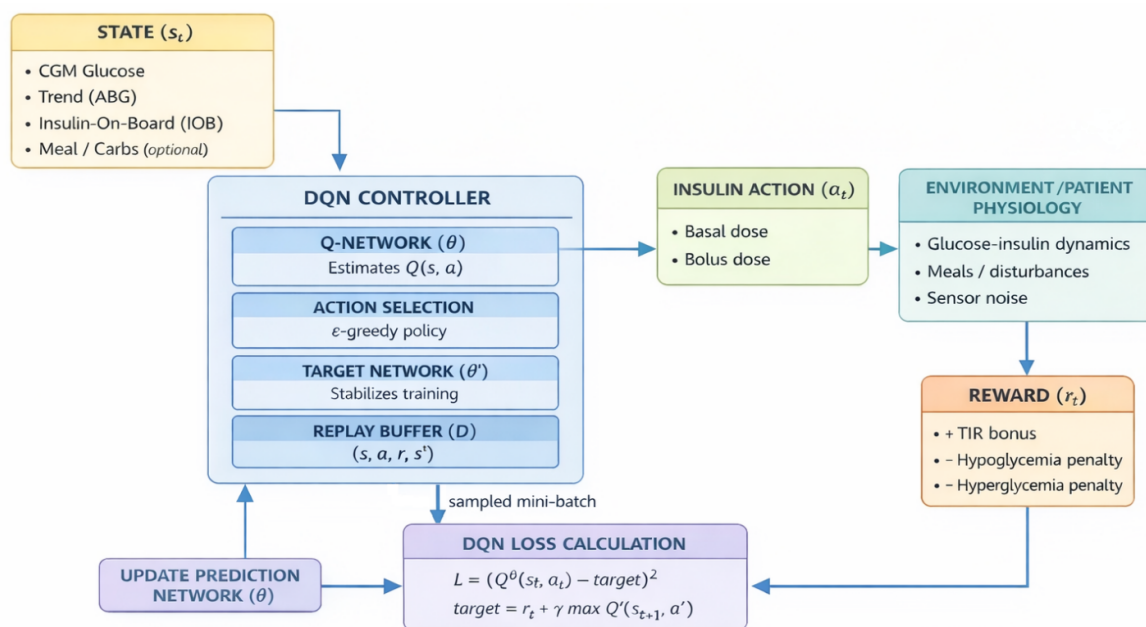


Figure 6. Deep Q-Networks (DQN)-based reinforcement learning framework for automated insulin delivery. Image created by the authors.

3.3.1. Multi-step DQN for blood glucose control

Wang *et al.*⁴⁴ presented a study that revisits automated glucose control as a value-based RL problem and extends DQN to account for the delayed effects of insulin. Two key modifications are introduced: (i) an “effective action” feature that exponentially decays previous insulin doses, transforming the partially observed dosing problem into a learnable Markov Decision Process, and (ii) the use of multi-step returns combined with prioritized experience replay, allowing the agent to consider longer-term consequences and focus learning on high-TD-error transitions.

Using a modified Simglucose/UVA-Padova environment implemented within the OpenAI Gym framework, the proposed approach demonstrates faster convergence than an Effective-DQN baseline and improves time-in-range performance. The learned policy favors more minor, timely insulin adjustments over significant, reactive corrections.⁴⁴ These improvements are attributed to reduced bootstrapping bias from multi-step targets and to more efficient gradient updates enabled by prioritized experience replay.^{41,44}

3.3.2. Deep Q-learning for ICU insulin dosing and glucose control

Symeonidis *et al.*⁴⁵ formulate inpatient glycemic management in the intensive care unit (ICU) as a value-based RL problem and train a DQN using real-world electronic health record data from the MIMIC-III database.

The learned policy is evaluated against sliding-scale insulin protocols and conventional prediction models, including linear and logistic regression and random forests, to improve prediction accuracy and TIR while limiting hypoglycemia.

The underlying Markov Decision Process is designed using: (i) a similar-patient transition model that selects the k-most comparable patients, (ii) an ICU-specific reward function with a strict target glucose range of 110–180 mg/dL centered near 145 mg/dL, and (iii) an ϵ -greedy behavior policy. From an initial cohort of 5,659 ICU patients and 65,242 records, 2,493 patients with sufficient historical data were retained for analysis. Performance is evaluated using MAE, RMSE, and TIR metrics.

The results show insulin-dose prediction errors of MAE and RMSE of 1.99 and 2.27, respectively, outperforming linear regression baselines (MAE 2.04, RMSE 2.43). Glucose prediction performance achieves an MAE of 12.16 mg/dL, RMSE of 15.42 mg/dL, and a TIR of 90.79%, compared to approximately 86–87% for baseline methods. Although some model-to-real comparisons are not significant after false discovery rate correction, pairwise improvements over linear regression remain statistically substantial ($p < 0.05$). Noted limitations include the absence of explicit meal modeling and insulin pharmacokinetics. The authors highlight the need for uncertainty calibration and prospective validation, and demonstrate the feasibility of applying deep reinforcement learning to real ICU electronic health record data using a clinically informed

reward design.⁴⁵ A comparison of representative offline RL approaches is provided in Table 2, illustrating how different algorithms balance safety and control performance.

3.3.3. Offline reinforcement learning on real-world clinical data

Beolet *et al.*⁴⁶ investigated offline RL for glycemia control using real-world data collected from the commercial DBLG1 closed-loop system. The dataset includes 100 patients and approximately 6.9 million transitions, derived from routine clinical usage rather than simulator-generated trajectories. The study evaluates BCQ, CQL, and TD3-BC algorithms under an offline learning setting.

Among the tested methods, TD3-BC achieved the strongest results. Compared to the behavior policy, the RL agent improved time-in-range TIR from 69.89% to 74.38%, reduced time-below-range from 3.59% to 2.73%, and lowered mean glucose from 156.6 mg/dL to 148.6 mg/dL. In scenarios with unannounced meals, the policy further increased TIR by approximately 8% and reduced mean glucose by 13.2 mg/dL relative to baseline control.

The study also applies Fitted Q Evaluation (FQE) to estimate clinical metrics such as TIR, time-below-range, and time-above-range directly from offline data, without relying on a simulator. In addition, a patient-wise personalization procedure using approximately three months of individual data resulted in average TIR improvements of about 1% and improved worst-case patient performance.

This work demonstrates that offline reinforcement learning can be trained and evaluated on real patient data

while improving clinically relevant glucose metrics. It reduces reliance on simulated environments and provides evidence of policy robustness under real-world variability.

Taken together, value-based RL studies for glucose control show encouraging results across both simulator and real-world settings, but comparisons need to be made carefully. For example, Wang *et al.*⁴⁴ test a multi-step DQN in a modified Simglucose/UVA–Padova simulator and report improved time-in-range under outpatient-style conditions with delayed insulin dynamics. In contrast, Symeonidis *et al.*⁴⁵ apply Deep Q-Learning to real ICU data from the MIMIC-III database using a stricter inpatient target range of 110–180 mg/dL. In their cohort of 2,493 ICU patients, the model achieved a TIR of 90.79% and slightly lower insulin-dose prediction errors (MAE 1.99, RMSE 2.27) compared to linear regression (MAE 2.04, RMSE 2.43).

Beolet *et al.*⁴⁶ also worked with real clinical data, but in an offline setting, where the Q-learning model is trained only on previously collected patient records. Their results show improved glucose stability compared to standard rule-based approaches, while keeping decisions within defined safety limits. Still, ICU glucose control is very different from outpatient closed-loop use—meals are controlled, patients are monitored continuously, and interventions are supervised. In addition, simulator studies use virtual cohorts, while the ICU study includes 2,493 real patients. Reward definitions and training setups also vary. Because of these differences, gains in TIR or prediction error across value-based RL studies should be interpreted within their specific clinical context rather than viewed as directly comparable performance numbers.

Table 2. Comparison of offline RL methods for safe glucose control

Method	How it stays safe	Data used	Why it works for glucose control
BCQ ³³	Only chooses actions similar to past data	Logged CGM and insulin records	Avoids insulin doses that were never seen before
CQL ³⁴	Learns conservative value estimates	Offline data with distribution shift	Reduces risky decisions that could cause hypoglycemia
TD3-BC ³⁵	Stays close to recorded behavior	Moderate-quality logged data	Improves control without drifting far from safe actions
PAINT ³⁶	Follows a safety-aware policy prior	Logged data plus preference input	Adjusts insulin delivery based on patient safety preferences
Multi-Step DQN ⁴³	Uses insulin history and longer return horizons	Simulated glucose trajectories	Better handles delayed insulin effects without aggressive dosing
ICU Deep Q-Learning ⁴⁴	Clinically shaped rewards and conservative policies	Real ICU EHR data	Shows value-based RL can safely learn insulin strategies from hospital data.

Abbreviations: EHR: Electronic health record; ICU: Intensive care unit; RL: Reinforcement learning.

3.4. CHO-agnostic, modular reinforcement learning

Carbohydrate-agnostic (“CHO-agnostic”) controllers remove the need for carbohydrate counting and meal announcements by learning insulin dosing directly from CGM and insulin histories. A modular reinforcement learning perspective makes this approach practical by separating the system into three components: (i) a sensing and feature-extraction module that windows and normalizes CGM data, (ii) a policy or critic network that proposes insulin actions, and (iii) a safety wrapper that can scale or veto unsafe actions. This modular separation reduces patient burden and limits brittleness caused by misestimated meals, while allowing each component to be improved independently, such as enhanced short-horizon prediction, alternative policy architectures, or stricter safety constraints.

From a clinical perspective, these announcement-free, modular designs are attractive because they reduce patient burden; however, they also place greater emphasis on robust safety filtering to compensate for the absence of explicit meal information.

Early studies demonstrated that fully closed-loop, announcement-free glucose control using deep RL is feasible *in silico*.⁴⁷ More recent modular approaches explicitly integrate glucose prediction and short-horizon planning around PPO-style policies to maintain CHO-agnostic operation.⁴⁷ Bioinspired controller designs have similarly advanced toward fully automated insulin dosing without requiring manual inputs.⁴⁸

3.4.1. G2P2C: A modular reinforcement learning algorithm for glucose control by prediction and planning in T1DM

The work presented by Hettiarachchi *et al.*⁴⁷ proposes a modular deep reinforcement learning controller that eliminates carbohydrate estimation and meal announcements. The agent relies solely on recent CGM and insulin histories as inputs and on insulin delivery rates as outputs. Built on a PPO framework, the approach introduces two plug-in components: a model-learning phase that fits a glucose dynamics model, and a short-horizon planning phase that rolls the learned model forward to refine immediate actions. Throughout training and deployment, insulin actions are constrained within physiological limits, and optimization is guided by a risk-index-based reward function that penalizes hypoglycemia.

In silico evaluations are conducted using a demanding protocol with approximately 180 g of daily carbohydrate intake across 20 virtual subjects (10 adults and 10 adolescents) in an FDA-accepted Type 1 Diabetes simulator. Performance is compared against state-of-the-

art RL methods and standard clinical basal–bolus strategies. G2P2C outperforms vanilla PPO in 18 out of 20 subjects, reduces failure rates, and achieves clinically meaningful results despite operating without meal announcements. Reported outcomes include adult TIR values in the low-to-mid 70% range and adolescent performance in the low-to-mid 60% range.⁴⁷

3.4.2. Bioinspired reinforcement learning toward fully automated, CHO-agnostic control

Lee *et al.*²⁵ propose a bioinspired reinforcement learning controller aimed at achieving fully automated AID operation without carbohydrate entries or meal announcements. The controller design incorporates a temporal homeostasis objective within the reward function to promote stable euglycemia, while personalizing temporal discounting to reflect individual pharmacological characteristics. This design choice guides the policy toward safe insulin dosing over clinically relevant time horizons.

In silico evaluations are conducted in the FDA-accepted UVA/Padova simulator (Section 2.2) using a demanding protocol, such as variations in insulin sensitivity and the dawn phenomenon, the trained policy demonstrates fully automated regulation during both basal and postprandial phases. In single-meal evaluations, the approach reports mean glucose levels of approximately 125 mg/dL, with approximately 90% of the time within the target range. The study further applies layer-wise relevance propagation to enhance interpretability, demonstrating robustness to sensor noise and avoiding insulin stacking, thereby reinforcing clinical plausibility. Overall, this work shows that carefully designed reinforcement learning objectives, combined with individualized temporal weighting, can achieve CHO-agnostic glucose control *in silico* and integrate naturally with modular architectures that include planning or safety layers, without reintroducing meal inputs

3.4.3. Explainable reinforcement learning for glucose monitoring

Adjevi *et al.*⁴⁸ proposed an explainable RL framework for glucose monitoring that combines signal preprocessing, CNN-based glucose prediction, SHAP value attribution, and an Actor–Critic RL agent. Exponential decay models preprocess insulin and related inputs, and a compact CNN predicts short-term glucose levels. SHAP values are used to identify the most influential CGM and insulin features, and these attributions guide a bounded RL adjustment step.

The method was evaluated on the OhioT1DM dataset. More than 98% of predictions fall within Clarke Error Grid

Zones A and B, indicating clinically acceptable accuracy. The model remains compact and trains in under one minute per patient for 1000 epochs. The RL component corrects near-target glucose deviations in approximately 25% of cases within 10 minutes, although larger deviations cannot always be fully restored within that time window.

Although developed as a research-only decision-support framework, the study shows that interpretability can be integrated directly into an RL structure without reducing predictive performance. This design direction is relevant for future glucose regulation systems that require both transparency and adaptive decision support.

Taken together, carbohydrate-agnostic and multi-agent reinforcement learning approaches demonstrate meaningful progress toward reducing patient burden while maintaining glycemic safety; however, cross-study comparison must be interpreted within the context of evaluation design. For example, G2P2C proposed by Hettiarachchi *et al.*⁴⁷ is evaluated under a demanding protocol with approximately 180 g/day carbohydrate intake across 20 virtual subjects, reporting adult TIR values in the low-to-mid 70% range and adolescent performance in the low-to-mid 60% range. In contrast, the bioinspired controller developed by Lee *et al.*²⁵ reports approximately 90% TIR in single-meal evaluations with mean glucose near 125 mg/dL under unannounced disturbances in the FDA-accepted UVA/Padova simulator. Similarly, the multi-agent basal-bolus advisor presented by Jaloli *et al.*⁴⁹ reports TIR values ranging from 78% to 94% across simulated scenarios, alongside marked reductions in hypoglycemia. Although these results are encouraging, differences in cohort composition, meal protocols, disturbance severity, simulation horizon, and target definitions limit direct numerical comparison. Because all three approaches are evaluated entirely *in silico*, the reported performance gains reflect simulator-specific dynamics rather than real-world behavioral variability or device-related uncertainty. Consequently, improvements across CHO-agnostic and multi-agent frameworks should be interpreted relative to protocol complexity and validation setting rather than as directly comparable benchmarks.

3.5. Multi-agent and hierarchical reinforcement learning

Multi-agent and hierarchical reinforcement learning approaches decompose the automated insulin delivery task into multiple cooperating decision-makers that reflect insulin pharmacology and clinical practice. A standard design assigns one agent, or low-level controller, to regulate basal insulin, while a second agent handles bolus dosing. These agents are coordinated using shared rewards or a

simple supervisory control logic. This separation lets each policy focus on a specific role, with one managing slow background glucose stabilization and the other responding to faster, meal-related changes. It also makes the system easier to modify, since one agent can be updated or replaced without retraining the entire controller, and it fits naturally with safety wrappers and scheduling constraints used in closed-loop systems. Recent reviews of reinforcement learning for diabetes management indicate that this type of decomposition is a practical design choice.⁵⁰

3.5.1. Basal-bolus advisor via multi-agent reinforcement learning

Jaloli *et al.*⁴⁹ present a multi-agent basal-bolus advisor based on soft actor-critic reinforcement learning, integrated with a blood glucose metabolic model and evaluated entirely *in silico* for multiple daily injection therapy. Their approach trains two continuous-control agents: a basal agent responsible for background insulin delivery and a bolus agent that adjusts insulin around meals. Together, these agents operate as a cooperative closed-loop controller.

The method is evaluated across three scenarios, including meal disturbances and changes in insulin sensitivity. Compared with conventional therapy, the proposed controller shows improved glycemic performance. Reported results include higher TIR (70–180 mg/dL), ranging from 78% to 94% across scenarios, along with an apparent reduction in hypoglycemia, from approximately 14–17% to 2–4%. Daily basal insulin usage is also reduced, and Control-Variability Grid outcomes are concentrated in the desired region. Overall, this study shows that separating basal and bolus decisions across coordinated agents can reduce variability, improve robustness to disturbances, and provide a flexible structure for adding safety constraints or future hierarchical extensions. This basal-bolus decomposition aligns with the multi-agent architecture illustrated in Figure 7, where coordinated basal and bolus agents share state and reward information to jointly regulate insulin delivery in a closed-loop manner.

3.6. Limitations of RL

Although RL has shown considerable promise for adaptive insulin control in T1DM management, several limitations restrict its translation from simulation to clinical use. The first primary concern involves safety and reliability. RL algorithms inherently require exploratory interaction to optimize insulin-dosing policies, which cannot be conducted safely on human subjects due to the risks of hypoglycemia and hyperglycemia. Consequently, most controllers are trained and validated in *in silico* environments such as the UVA/Padova and

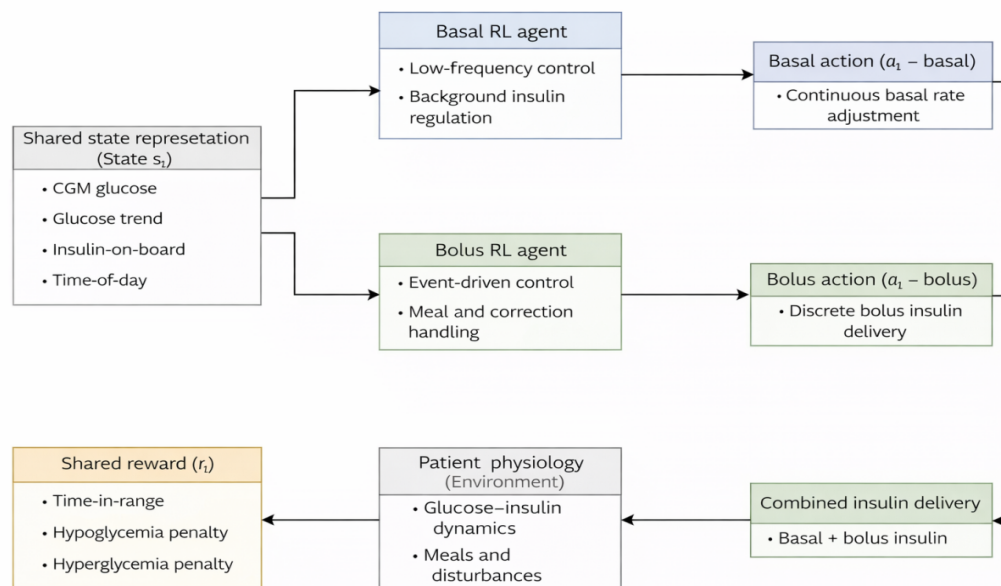


Figure 7. Basal–bolus insulin advisor implemented via a multi-agent reinforcement learning framework with shared state and reward representations. Image created by the authors.

Abbreviation: CGM: Continuous glucose monitoring; RL: Reinforcement learning.

Although physiologically detailed, Simglucose simulators fail to fully capture inter-patient variability, behavioral fluctuations, and unexpected disturbances encountered in daily life.^{17,21,27,30,35} This dependence on virtual simulations contributes to a simulation-to-reality performance gap, where policies that achieve stable glucose regulation *in silico* often degrade under real-world conditions.

A second limitation is data dependency and generalization. Offline RL methods, such as BCQ, CQL, and TD3-BC, mitigate safety risks by learning exclusively from pre-collected glucose–insulin trajectories.^{33–36} However, their effectiveness depends heavily on the quantity, quality, and representativeness of the data. When faced with limited or biased datasets, these agents may overestimate value functions or produce unsafe actions in out-of-distribution states.^{33,34} In addition, achieving personalization remains challenging: each patient exhibits unique insulin sensitivity, dietary behavior, and physical activity patterns that evolve over time. Standard RL agents lack efficient mechanisms for adapting to non-stationary, individualized glucose dynamics, often requiring manual recalibration or retraining.^{37,46,47}

A third challenge relates to interpretability and clinical validation. Deep RL policies, while effective in simulation, operate as black-box models that provide little transparency into decision-making processes, thereby limiting clinician

trust and regulatory acceptance.^{29,44} Furthermore, computational complexity remains a practical barrier. Policy-gradient and actor–critic architectures (e.g., PPO, TD3-BC, DQN) require substantial computational resources and memory, which are challenging to implement in embedded insulin pumps or wearable glucose sensors with strict energy and latency constraints.^{28,36} These factors collectively restrict the scalability of RL-based control systems in real-world healthcare environments.

An additional concern relates to dependence on fixed *in silico* environments, particularly the widely used FDA-accepted UVA/Padova simulator. Although this platform provides physiologically grounded and reproducible preclinical evaluation, it remains a parameterized mathematical model governed by structured glucose–insulin dynamics and predefined variability distributions. When reinforcement learning controllers⁵¹ are both trained and benchmarked within closely related simulation frameworks, there is a risk that policies may partially adapt to simulator-specific regularities rather than universally robust physiological behavior. In continuous-control settings, optimization procedures may exploit implicit assumptions embedded in compartmental formulations or disturbance protocols, a pattern consistent with broader findings in real-world reinforcement learning and offline RL under distribution shift.^{52,53} Consequently, improvements reported exclusively *in silico* may reflect alignment with simulator artifacts rather than resilience to

real-world behavioral variability, sensor noise, and device constraints. Recent reviews of RL-based automated insulin delivery systems further emphasize the need for broader validation beyond simulation environments.⁵⁴

A further concern that cuts across the reviewed RL literature is the lack of reproducibility and independent validation. Reward functions, state representations, and training configurations are inconsistently reported across studies, making direct replication difficult. More critically, almost no study evaluates its controller on a truly independent cohort. In the UVA/Padova simulator, virtual patients are generated from the same parameter distributions used during training, so evaluation is not genuinely independent of the training environment. Among the few studies using real data, Beolet *et al.*⁴⁶ validated TD3-BC on 100 patients from the DBLG1 system and Symeonidis *et al.*⁴⁵ tested on 2,493 ICU patients from MIMIC-III, yet neither reported external validation on a separate institutional cohort. Establishing standardized reporting guidelines and independent benchmarking protocols would be an important step toward more credible and reproducible RL-based AID research.

In contrast, FL offers complementary strengths that address several of these constraints. While RL focuses on optimizing sequential decision-making for glucose control, FL emphasizes privacy-preserving and distributed model training across multiple institutions or devices.⁵⁵⁻⁶⁰ By allowing collaborative learning without centralized data sharing, FL can capture diverse physiological and behavioral patterns, thereby improving model generalization and personalization across heterogeneous patient populations. Nevertheless, FL alone cannot manage real-time insulin control. Consequently, integrating RL with FL has emerged as a promising pathway toward developing adaptive, privacy-preserving, and clinically reliable artificial-pancreas systems capable of large-scale deployment.

These limitations point to the need for learning frameworks that can leverage larger, more diverse patient populations without centralizing sensitive data. Federated learning, therefore, represents a complementary approach, enabling collaborative model development while supporting personalization and privacy in glucose-related applications.

4. Federated learning

Federated learning (FL) was initially proposed by McMahan *et al.*⁵⁵ as a decentralized learning framework that enables collaborative model training without requiring direct access to sensitive local data. In healthcare applications such as glucose monitoring, FL enables multiple clinical

institutions or personal devices to jointly train predictive models by sharing model updates rather than raw patient data.⁵⁶ Electronic health records are a significant source of information for clinical machine learning tasks⁵⁷; however, models trained on data from a single hospital often exhibit population-specific bias. At the same time, privacy regulations and ethical constraints limit direct data sharing across institutions. FL addresses these limitations by enabling collaborative model development while ensuring that patient data remains stored locally.

In the following sections, we focus on federated learning mechanisms that directly influence personalization, robustness, privacy, and scalability in glucose-related tasks, while drawing on non-glucose studies only where they provide transferable system-level insights. Depending on how data is distributed across participating clients, FL is commonly categorized into three settings: horizontal FL, vertical FL, and federated transfer learning. Horizontal FL applies when different clients share similar feature spaces but correspond to distinct patient populations. Vertical FL, in contrast, is used when institutions observe the same individuals but collect different types of features. Federated transfer learning targets scenarios with limited overlap in both users and features, where knowledge from pre-trained models is transferred and adapted using additional information.⁵⁸

4.1. Federated learning for glucose

Federated learning has attracted growing interest in medical prediction tasks because it enables collaborative model training without sharing patient-level data. This is particularly important in healthcare, where privacy regulations such as HIPAA and GDPR restrict how sensitive patient information can be stored and transferred.⁵⁹ By aggregating knowledge across distributed clients, FL reduces dependence on centralized datasets and can improve model generalization across different patient populations and clinical environments.⁶⁰

Glucose-related healthcare data are naturally heterogeneous. Hospitals and individual patients rely on different glucose sensors, insulin delivery systems, clinical protocols, and data sampling schedules. As a result, the collected data are often non-identically distributed (non-IID), which poses challenges for traditional machine learning methods that typically assume similar data distributions across training samples.⁶¹ In federated settings, this heterogeneity can negatively affect training stability and overall model performance.

Several FL strategies have been proposed to address performance degradation caused by non-IID data. Some methods introduce shared representation spaces or

filtering mechanisms that enable clients with different data characteristics to collaborate more effectively.⁶² Other studies focus on personalization strategies and client clustering approaches that aim to mitigate heterogeneity by enabling models to adapt to local patient data while still benefiting from global training.⁶³

In addition, CGM and insulin pump datasets are often large and expensive to transmit across institutions. This makes centralized learning inefficient from both computational and logistical perspectives. FL alleviates these limitations by exchanging model parameters or updates rather than raw data, thereby reducing communication costs, improving scalability, and supporting deployment in environments with limited bandwidth or infrastructure.⁶⁴ Figure 8 depicts the standard FL architecture for training glucose-related models, in which distributed clients perform local updates on private CGM and insulin data and share only encrypted model parameters with a central aggregation server to produce a global model.

In glucose prediction, this system provides a natural fit for glucose prediction and diabetes management because data are inherently distributed across patients, devices (CGMs, insulin pumps), and clinical systems. As shown in Figure 8, the standard FL architecture relies on a central server (or orchestrator) that coordinates training across multiple edge devices or institutions. The global model parameters W_t are distributed from the server to participating clients, who perform local training on their private data. Rather than transmitting raw data, each client uploads locally computed model updates W_{t+1} to the server. The server aggregates these updates, commonly via weighted averaging, and produces an updated global

model, which is redistributed in the following training round. This iterative process enables scalable learning while preserving data locality and offering privacy advantages relative to centralized training.

In this review, we focus on FL methods relevant to diabetes management and glucose-related tasks, synthesizing key developments in FL architectures, aggregation strategies, privacy and security mechanisms, and federated data analysis that directly impact personalized and privacy-preserving glucose modeling and decision support. For studies and works discussed outside the glucose conversation, we focus on the underlying system and privacy insights, which are directly applicable to glucose stimulation.

4.2. Architecture

From an architectural perspective, federated learning systems are generally divided into two categories: (i) centralized architectures and (ii) decentralized architectures. Centralized FL dominates current healthcare deployments due to its simplicity and coordination benefits, while decentralized FL is increasingly explored in settings where trust, scalability, and fault tolerance are primary concerns.

4.2.1. Centralized federated learning

Centralized FL is the most widely adopted architectural paradigm in existing literature. Foundational FL methods, such as FedAvg and its extension FedProx, follow a centralized server-based design.^{62,64} In this setting, a central server coordinates the training process by collecting local model updates from participating clients, aggregating

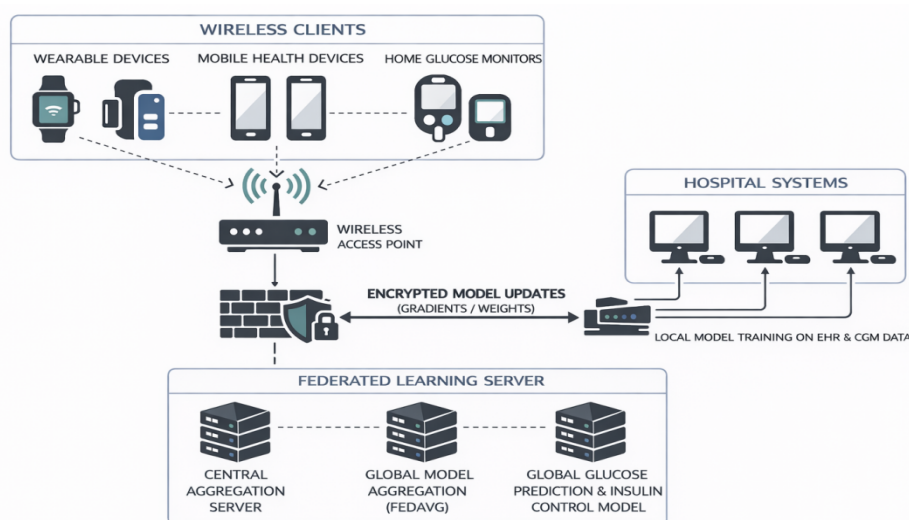


Figure 8. Standard federated learning architecture for collaborative model training. Image created by the authors.
Abbreviations: CGM: Continuous glucose monitoring; EHR: Electronic health record.

these updates, and redistributing the updated global model. Several architectural variants of centralized FL have been proposed to improve scalability, personalization, and resource efficiency, as summarized below.

Astillo *et al.*⁶⁵ proposed A deep learning-based anomaly detection system composed of estimation and classification models for a diabetes management control system (DMCS). The estimation model is used to forecast glucose levels, while the classification model is used to detect anomalous data points. Both independent learning and FL methods are employed, along with a post-quantization compression technique, to address data privacy concerns and reduce the computational demands of deep learning models. Experimental results show that FL-based models (horizontal centralized FL) achieve higher recall rates of approximately 98.6% compared to independent learning-based models. Furthermore, the FL-aided CNN model achieves an average recall rate of 99.24%, which is significantly higher than that of other FL-aided models. In conclusion, the system employs a client-server-based, centralized FL architecture, in which locally trained CNNs and other deep learning models are aggregated into a global model.

Other works outside the glucose context, such as Jaramillo-Velez *et al.*⁶⁶ who have proposed a novel model aggregation algorithm for federated learning has been proposed to mitigate the risk of exposing sensitive data. Secure aggregation via multiparty computation has been proposed; however, it can only be achieved through uniform aggregation of masked local models, which poses a problem when working with real-valued data, as there is no measure on the reals that is invariant under the masking operation, and hence information leakage is inevitable.

To address this issue, the authors proposed a secure parameter aggregation method that employs a torus rather than a finite field to perform data shifting. This ensures a uniform distribution on the torus while avoiding loss of accuracy. Experimental results conducted on the MNIST and CIFAR-10 datasets show that the torus-based approach outperforms finite-field secure aggregation while achieving the same accuracy and cosine similarity to non-secure aggregation algorithms. Privacy is preserved by demonstrating that the mutual information between the local model parameters and the models shared with the server is zero, ensuring no information leakage.

Hence, the torus-based protocol is a safe choice that provides perfect privacy while easily avoiding overflow issues, thereby proving an effective solution for centralized cross-silo federated learning. This work follows a centralized FL paradigm, as all coordination, aggregation, and global model updates are performed by a single central

server.

Another recent work on centralized FL, Xu *et al.*⁶⁷ proposed a model to tackle information leakage caused by the interchange of gradients. Although secure aggregation is integrated, recent mechanisms are vulnerable to disaggregation attacks. The authors propose TAPFed, which employs a threshold functional encryption scheme that allows a limited number of aggregators.

Before training begins, TAPFed assigns cryptographic keys, fixed fusion weights, and labels for each training round. In every round, clients locally train their models and encrypt their updates using MCFE, which are then sent to multiple aggregators that directly perform aggregation in the ciphertext domain without learning the aggregated model update. Once a sufficient number of aggregated ciphertexts exceeds the threshold, clients jointly decrypt the global update and proceed to the next training round. Experimental results, supported by statistical analysis across several baselines, show that TAPFed achieves equivalent model performance while reducing transmission overhead by approximately 29–45%.

4.2.1.1. Hierarchical federated learning

Hierarchical FL (HFL) extends the classical client-server model by introducing intermediate aggregation layers between clients and the central server. In this architecture, groups of clients communicate with nearby edge servers, which perform partial aggregation of local updates before transmitting aggregated results to a higher-level server. This design reduces communication overhead, improves scalability, and is well-suited to geographically distributed environments.^{68,69} After one or more rounds of edge-level aggregation, intermediate models are forwarded to the cloud for final global aggregation.

Singh *et al.*⁷⁰ proposed a hierarchical federated learning model for large, integrated healthcare organizations that include several institutions. There are two levels of aggregation: the first collects local updates at regional centers and forwards them to the second level for aggregation, where global updates are formed. After training and updating the global model, the updated model is returned to the local nodes for additional rounds of training and updating. This reduces message passing traffic and improves the scalability of the HFL structure.

The HFL structure also includes a novel Aggregated Gradient Perturbation Mechanism that alleviates noise in model updates while preserving both privacy and utility. The performance is evaluated on real healthcare and synthetic datasets, and the results show that the HFL model outperforms other FL models, including FLBM-IoT and PPFLB. The proposed model achieves 97% accuracy and

exhibits 30% less information leakage, thereby confirming that the HFL model is a highly secure approach for healthcare systems. It presents a safe, flexible, and effective model that is readily adaptable to diverse healthcare settings (including glucose-related applications), improving workflow efficiency while enhancing patient care without compromising privacy.

4.2.1.2. Personalized federated learning

Personalized FL targets scenarios where a single global model does not perform equally well across all clients due to substantial differences in local data distributions. In glucose-related healthcare applications, strong non-IID characteristics naturally arise from variations across patients, hospitals, sensing devices, and individual physiological responses. As a result, purely global models often fail to capture patient-specific dynamics.

Personalized FL approaches typically begin with a shared global model and then allow each client to adapt selected components locally. For example, Arivazhagan *et al.*⁷¹ proposed personalization layers in which specific parameters are locally trained, while shared layers are aggregated across clients. This design supports individualized model adaptation while still preserving collaborative benefits across distributed participants.

Yang *et al.*⁷² proposed the use of personalized federated deep learning models trained across multiple Internet of Things (IoT) devices, while keeping sensitive and personal data on the devices themselves and avoiding data centralization. A personalization component is incorporated into each FL model, enabling adaptation to individual users. This design reflects the fact that insulin sensitivity, food preferences, and physical activity affect glucose levels differently across individuals. In addition, the model is optimized for the limited processing power, memory, and battery life of IoT devices. Overall, a user-specific personalized FL model is used to support individualized glucose prediction without sharing sensitive data.

Yang *et al.*⁷³ have used a personalized clustering-based federated deep learning model to address challenges related to patient variability and to optimize FL for glucose prediction. Tailored models are developed by clustering patients based on carbohydrate (CHO) intake, using Simple Recurrent Neural Network (SimpleRNN) and Gated Recurrent Unit (GRU) architectures. The results show that, for local patients who contribute to training both cluster and global models, the Clu-FDL approach achieves high precision (0.93), recall (0.96), and F1 score (0.95), along with low RMSE values (11.08 ± 1.77 mg/dL). In addition, the model outperforms non-clustering

approaches in stability for new patients. This demonstrates how clustering-based personalization can be integrated into FL to support personalized glucose prediction while preserving user data privacy.

Zhang *et al.*⁷⁴ proposed a federated meta-learning framework with fine-grained classifier reconstruction (FedFGCR) to address performance degradation under strongly non-IID data distributions. The method uses a Model-Agnostic Meta-Learning (MAML)-inspired adaptive interpolation strategy to combine global and local feature extractors before local training, and introduces a regularization term to incorporate global semantic information into client models. In addition, classifiers are reconstructed at a fine-grained level to improve personalization rather than relying on coarse global updates. Experimental results show accuracy gains ranging from 0.96% to 5.41% on the CWRU dataset and from 5.01% to 9.95% on the KAIST dataset compared to mainstream FL methods. Although developed for fault diagnosis, the personalization mechanism is applicable to glucose-related healthcare, where patient-specific differences in insulin sensitivity and glucose dynamics create heterogeneous data distributions. Such meta-learning-based adaptation can improve individualized glucose prediction or control while preserving privacy.

Oh *et al.*⁷⁵ introduced a personalized federated learning framework called FedBABU. In the proposed model, the entire network is split into the body (feature extractor), which is related to universality, and the head (classifier), which is related to personalization. In this method, the feature extractor is trained collaboratively across clients, while the final classification layer remains local and is updated individually using each client's own data. This design helps reduce the impact of heterogeneous data distributions and prevents the global model from overwriting client-specific patterns. Experimental results show that FedBABU achieves consistent accuracy improvements of approximately 2–6% over standard FedAvg under non-IID conditions, along with more stable convergence behavior. Although originally evaluated on image classification tasks, the separation between shared representations and local decision layers is applicable to glucose-related healthcare applications. In CGM systems, a shared temporal feature extractor can learn common physiological patterns across patients, while patient-specific output layers can adapt to differences in insulin sensitivity, meal response, and daily glucose variability. Such a structure supports individualized glucose prediction and potential federated insulin control while maintaining privacy-preserving collaboration.

Finally, Dinh *et al.*⁷⁶ proposed a personalized FL method

known as pFedMe, which balances global collaboration and local personalization through a regularization-based objective. In this framework, each client maintains a personalized model while remaining close to the global model, allowing adaptation to heterogeneous data distributions rather than enforcing a single shared solution. To evaluate performance, pFedMe was compared with FedAvg and Per-FedAvg under both strongly convex and nonconvex settings using MNIST and Synthetic datasets. For the MNIST dataset, the personalized model of pFedMe achieves 1.1–1.5% higher accuracy than its global model, Per-FedAvg, and FedAvg under strongly convex conditions, and 0.9–1.3% improvements under nonconvex settings. For the Synthetic dataset, the gains are more pronounced, reaching up to 6.1% compared to its global model and up to 5.2% over FedAvg. In addition, the personalized model shows consistently lower training loss and faster convergence than the baseline methods across different settings. Although evaluated on benchmark datasets, the regularization-based personalization mechanism is relevant to glucose-related healthcare applications, where differences in insulin sensitivity and glucose dynamics create heterogeneous data distributions. Such an approach can support individualized glucose prediction or federated insulin control while preserving privacy.

4.2.1.3. Split federated learning

Split FL differs from standard federated learning by dividing the neural network between the client and the server, such that neither side ever holds the complete model during training. Instead of transmitting full model updates, the client computes activations up to a predefined split layer and forwards these intermediate representations to the server, which then completes the remaining forward and backward computations.⁷⁷

This approach preserves data locality, reduces computation on constrained devices, and provides an additional layer of privacy because only encoded activations are exchanged; raw input features remain local and are never directly shared. At the same time, split learning allows multiple clients to participate in collaborative training without exposing their internal model structure. By assigning deeper, more computationally intensive layers to the server, this approach reduces computational and memory requirements on resource-constrained client devices, which is particularly relevant in healthcare and edge-based environments.

Alrashed *et al.*⁷⁸ proposed a privacy-preserving vertical FL framework that uses split learning to address challenges of machine learning models on vertically partitioned data. The model enables collaborative model training while

preserving patient privacy. This approach offers several advantages, such as enhanced data privacy, efficient communication, and adaptability to heterogeneous data distributions.

In the first stage of the model, each client trains the lower layers of the global model using only overlapping samples. Then, each client shares only the trained model's output, known as embeddings, with the server, rather than sharing complete model parameters or raw patient data, to preserve patient privacy. The server receives the embeddings and feeds them into the server-side top model, thereby completing the training process. It then makes predictions and computes the gradients required to update the global model. The server splits the computed gradients for each client and shares them individually, which further preserves privacy.

Split learning reduces communication overhead by avoiding the need to share complete model parameters. It also works well with vertically partitioned datasets, where healthcare providers hold different features. Three datasets were used for evaluation; in this work, the diabetes prediction dataset is considered and compared with a centralized learning baseline model, in which all data are integrated into a single repository to achieve optimal performance. The performance of the PPVFL-SplitNN model was comparable to the centralized learning baseline, achieving accuracies of 88.70 and 90.22, respectively, thus offering a practical solution for sensitive healthcare networks.

4.2.1.4. Graph federated learning

Graph FL extends the federated learning paradigm to settings where data are naturally represented as graphs. In this framework, each client maintains a local subgraph, while the learning objective is to capture patterns that emerge across the distributed graph as a whole. During training, only model parameters are exchanged, thereby preventing direct exposure of node attributes, edge information, or graph connectivity structures.⁶⁹ This design enables the learning of global graph representations without requiring complete graph centralization. It also supports heterogeneous graph partitions, where clients may control different node types or sparsely connected regions. As a result, graph FL is well-suited for applications such as sensor networks, biomedical graphs, and other structured healthcare data, where data sharing is restricted or impractical.⁷⁹

Piao *et al.*⁸⁰ proposed a graph-based explainable FL approach for predicting blood glucose levels. They developed a graph-attentive memory model that integrated a graph attention network with a Gated Recurrent Unit

(GRU). This hybrid model enhances blood glucose level prediction by modeling graph correlations, which address the lack of correlation among attributes in previous methods, and by capturing dynamic attribute relationships with attention weights that evaluate the significance of attributes over time. Furthermore, federated learning is employed to facilitate secure analysis and population pattern analysis. The proposed approach was validated on the OhioT1DM'18⁸¹ and OhioT1DM'20⁸² datasets, comprising 12 participants, and focused on six key attributes.

4.2.1.5. Asynchronous federated learning

Asynchronous FL eliminates the requirement that all clients train and upload updates simultaneously. Instead, each client sends its update upon completion of local training, and the server incorporates it immediately rather than waiting for slower participants. This design helps avoid delays caused by stragglers, shortens overall training time, and supports devices or hospitals with intermittent connectivity or limited computational capacity.

Allowing updates to arrive continuously enables ongoing model improvement and makes asynchronous FL suitable for real-world deployments with uneven client workloads. However, this flexibility introduces challenges related to model staleness, where older updates may negatively affect convergence. Addressing this issue typically requires aggregation or optimization strategies that explicitly account for update delays in asynchronous FL systems.⁸³

In Piao *et al.*⁸⁴, the cold-start problem has been addressed, which arises from insufficient blood glucose data in CGM. To handle this, they propose GluADFL, an asynchronous decentralized federated learning framework for blood glucose level prediction. Eight baselines were evaluated across four distinct T1DM datasets comprising 298 participants, and the approach demonstrated superior performance in predicting blood glucose levels. Patient data are stored across various communication networks, ranging from interconnected to more structured settings. The asynchronous framework supports flexible participation by accommodating inactive patients, even when up to 70% of clients are inactive. Their work differs from standard client-server FL by enabling decentralized training that better addresses scenarios in which patients do not interact consistently, and network connectivity varies, while preserving data privacy.

4.2.2. Decentralized federated learning

Decentralized FL reduces reliance on a central coordinator by adopting distributed computing paradigms. Examples

include peer-to-peer networks.

Kalra *et al.*⁸⁵ proposed a decentralized FL model called ProxyFL, where each participant maintains two models: a private model and a publicly shared proxy model designed to protect user privacy. This design allows for efficient communication without directly exposing private model parameters. Results on several healthcare-related datasets show that the ProxyFL protocol outperforms existing alternatives while achieving lower communication overhead and stronger privacy guarantees.

The experiments were conducted using several benchmark datasets, including MNIST, Fashion-MNIST, and CIFAR-10. The experimental setup was executed on a server equipped with eight V100 GPUs, corresponding to eight clients. In each case, every client was assigned 1k samples for MNIST and Fashion-MNIST or 3k samples for CIFAR-10, drawn from non-overlapping private subsets of the training data. This setup increased the difficulty of the classification task, as each client had access to only a subset of the overall training data.

For each user, a randomly chosen class was assigned, and a fraction of the client's private data was sampled from that class, while the remaining data was drawn from all other classes in an IID manner. As a result, clients were required to learn collaboratively to generalize effectively to the IID test set. The results show that centralized schemes such as FedAvg and FML-proxy exhibit limited learning progress compared to their decentralized counterparts, AvgPush and ProxyFL-proxy. In addition, ProxyFL-private consistently outperforms FML-private during training, indicating that ProxyFL provides a stronger learning signal to the private model than FML.

Tripathy *et al.*⁸⁶ proposed FedHealthFog, which was developed to overcome resource-constrained distributed learning in IoT-enabled healthcare systems. Conventional FL methods face significant challenges due to substantial computational requirements and high communication costs, which often lead to incomplete training. The proposed method addresses these challenges by elevating selected fog nodes to local aggregators within the federated learning architecture.

A sophisticated heuristic greedy technique is then employed to optimize the selection of the fog node as the global aggregator in each communication cycle between edge devices and the cloud. The results show a significant reduction in communication latency and energy consumption across three algorithms evaluated in the study. This positions FedHealthFog as a promising approach for supporting delay-sensitive applications in resource-constrained IoT healthcare environments.

Liang *et al.*⁸⁷ have explored a decentralized blockchain-based FL approach. This work attempts to address bias in healthcare data by improving fairness in predictive models using FL and by adding trust through a blockchain-assisted FL setup. The main objective is to develop a decentralized FL platform that improves fairness in healthcare prediction tasks while maintaining patient data privacy.

The authors have used a blockchain platform to support machine learning models trained on existing healthcare data stored off-chain. Their design follows a two-cycle training process in which fairness metrics are incorporated directly into FL and enforced via a blockchain consensus mechanism. This allows the global model to provide feedback that is then used to update the local training models. The system is implemented using Aplos smart contracts, microservices, the Rahasak blockchain, and Apache Cassandra-based distributed storage.

In a pilot study across five simulated medical centers, the model achieves good prediction accuracy while maintaining fairness, with only reasonable overhead. Overall, these decentralized architectures aim to enhance privacy, scalability, and fault tolerance and are increasingly relevant for large-scale collaboration among hospitals, IoT devices, and edge systems.⁷⁸ To consolidate these architectural choices, Table 3 compares major FL paradigms and illustrates how each supports secure and scalable glucose modeling.

4.3. Aggregation

In a typical federated learning setup, a central server maintains the global model and updates it using information received from participating clients. This update step is commonly referred to as aggregation. The most widely used aggregation method is FedAvg, introduced by McMahan *et al.*⁴⁸ In FedAvg, client updates are aggregated via a weighted average, with weights often proportional to each client's local dataset size.

At the beginning of each training round, the server sends the current global model to the participating clients. Each client then performs local training for a limited number of epochs and returns an updated model to the server. These updates are collected and merged to form the next global model.

Although FedAvg is simple and widely adopted, it is well known to perform poorly in realistic settings. In particular, non-IID data distributions, unequal amounts of local computation, and unbalanced datasets across clients can significantly degrade performance. To address these limitations, several aggregation strategies have been proposed that modify or extend the FedAvg framework. One such method is FedProx⁸⁹, which allows clients to perform different amounts of local work depending on their computational resources. FedProx introduces a proximal regularization term into the local objective, preventing client models from drifting too far from the

Table 3. Comparison of federated learning architectures for privacy-preserving glucose control

Method	How it stays safe	Data used	Why it works for glucose control
CFL	Keeps raw data local; shares only model updates	CGM, insulin, IoT sensor data	Secure collaborative glucose modeling
PFL	Local adaptation on top of a shared global model	Patient-specific CGM and lifestyle data	Captures individual insulin sensitivity and behavioral differences
CPFL	Aggregates updates within similar patient groups	CGM and carbohydrate (CHO) intake data	Improves prediction stability for new or low-data patients
GFL	Exchanges parameters while preserving graph structure	Graph-based glucose, insulin, and meal features	Models physiological relationships without centralizing structured data
AFL	Accepts updates without waiting for all clients	CGM data with intermittent availability	Handles irregular monitoring and client dropouts common in real-world glucose data
ADFL	No strict synchronization; decentralized updates	Multi-source CGM data across heterogeneous networks	Addresses cold-start and high client inactivity while preserving privacy
PBDFL	Peer-to-peer proxy model sharing	Distributed healthcare model parameters	Removes the central coordinator and improves robustness in collaborative healthcare learning
BEFL	Distributed ledger for update verification	Encrypted healthcare model updates	Improves trust and auditability for cross-institution glucose data sharing

Abbreviations: FL: Federated Learning; CFL: Centralized FL; PFL: Personalized FL; CPFL: Clustered personalized FL; GFL: Graph FL; AFL: Asynchronous FL; ADFL: Asynchronous decentralized FL; PBDFL: Proxy-based decentralized FL; BEFL: Blockchain-enabled FL.

global reference model and improving convergence under heterogeneous conditions.

Other aggregation approaches focus on optimization behavior. FedNova addresses inconsistencies arising from uneven local training by normalizing client updates prior to aggregation.⁹⁰ This normalization reduces bias introduced by variable local computation. Another set of methods falls under the FedOpt framework⁹¹, which applies adaptive optimization techniques such as Adagrad, Yogi, or Adam at the server during aggregation. These methods have demonstrated improved convergence in non-IID settings. In practice, FedAdam often produces more stable updates, while FedYogi tends to yield smoother training behavior across rounds.

In the clustering-based federated learning framework⁶⁶, each client's contribution to the aggregated model is weighted by the proportion of data it holds, defined as $\frac{n_k}{n}$, where n_k represents the number of local samples at client k , and n is the total number of samples across all participating clients. Cluster-specific models are computed by aggregating updates from clients belonging to the same cluster using weighted averaging:

$$w_{cluster_i}^{t+1} = \sum_{k \in cluster_i} \frac{n_k}{n} w_k^{t+1} \quad (14)$$

Here, $w_{cluster_i}^{t+1}$ represents the model parameters for cluster i after the $(t+1)^{th}$ communication round, while w_k^{t+1} denotes the locally updated model from the client k . Each update is weighted by $\frac{n_k}{n}$. This weighting ensures that clients with larger datasets have a proportionally greater influence on the cluster model, leading to more representative learning within each cluster.

In parallel, a global model is obtained by aggregating updates from all clients:

$$w_{global}^{t+1} = \sum_{k=1}^K \frac{n_k}{n} w_k^{t+1} \quad (15)$$

In this expression, w_{global}^{t+1} denotes the updated global model after round $t+1$, obtained by combining the locally trained models w_k^{t+1} from all K clients. While cluster aggregation captures similarities among patient groups, the global model summarizes broader population trends and improves generalization across diverse data distributions.

This two-level aggregation strategy enables the model to capture both cluster-level patient similarities and overall population trends. Since training is performed locally on patient devices and only aggregated model parameters are shared, sensitive glucose data remain private while supporting accurate and personalized glucose prediction.

In the asynchronous, decentralized, FL setup⁷¹, each active node updates its model by averaging its own parameters with those received from neighboring nodes, without relying on a central server. The update is computed as:

$$\hat{w}_n^{t-1} = \frac{1}{|N_n^{t-1}| + 1} \left(\sum_{n' \in N_n^{t-1}} w_{n'}^{t-1} + w_n^{t-1} \right) \quad (16)$$

Here $\hat{w}_n^{t-1}$ is the updated model estimate at the node n after communication, is the node's local model before aggregation, and $w_{n'}^{t-1}$ represents the models received from neighboring nodes. The set N_n^{t-1} contains the neighbors that successfully exchanged information with the node n , and $|N_n^{t-1}|$ is the number of such neighbors. Averaging the received updates helps smooth the learning process and makes the system resilient to dropped connections or inactive nodes, while still preserving privacy since only model parameters are shared.

After aggregation, each node continues training using its local glucose data. Since only model parameters are exchanged and no raw data is shared, patient privacy is preserved while the decentralized design remains robust to inactive nodes and variable network connectivity.

In the FedGlu FL approach, each client trains a local model using its own CGM data and sends only the updated model parameters to a central server.⁹² The server combines these updates using weighted averaging, where each client's contribution depends on the amount of local data it holds. Clients with more glucose samples, therefore, have a stronger influence on the global model.

$$w_{t+1} = \sum_{k=1}^K \frac{n_k}{n} w_k^{t+1} \quad (17)$$

Here, w^{t+1} represents the global model after the $(t+1)$ -th training round, while w_k^{t+1} is the model obtained by the client k after local training. The total number of clients involved in the learning process is denoted by K . Each client's update is scaled by the fraction $\frac{n_k}{n}$. By assigning larger weights to clients with more data, this aggregation approach helps the global model better reflect overall population patterns while maintaining distributed training and privacy preservation.

This aggregation process allows learning across multiple patients without sharing raw CGM data. Since all glucose data remain on local devices and only model parameters are exchanged, patient privacy is preserved. At the same time, the global model benefits from population-level information, improving glucose prediction performance, especially for rare hypoglycemic and hyperglycemic events.

4.4. Privacy components

Federated learning systems involve many distributed components, which naturally increases the number of potential security and privacy risks. Although raw client data are never directly shared, privacy concerns still arise through the communication of model parameters and gradients. As a result, ensuring secure communication and robust aggregation mechanisms is critical in FL environments.

One of the most commonly discussed threats is data poisoning, also known as model poisoning. In these attacks, a malicious client intentionally submits manipulated data or crafted model updates in order to bias the global model. In more severe cases, poisoning may result from a server-side compromise or intrusion.⁹³ Another class of threats includes data leakage attacks, in which adversaries attempt to reconstruct sensitive training data by analyzing shared model parameters or updates.⁹⁴

Membership inference attacks pose an additional risk, allowing an attacker to determine whether a specific data record was included in a client's training dataset. Such attacks can be passive, in which the attacker only observes gradients or parameters, or active, in which the attacker directly participates in the training process.⁹⁵ Model inversion attacks go a step further by attempting to recover sensitive features from the trained global model itself.⁹⁶ Closely related are gradient inversion attacks, which aim to reconstruct the original training samples directly from shared gradients or model updates.⁹⁷

A broader overview of privacy protection strategies and their relevance to glucose-related FL is summarized in Table 4, which highlights commonly used mechanisms, including knowledge abstraction, secure aggregation, differential privacy, and robust aggregation defenses.

These attack vectors highlight the need for privacy-preserving techniques in FL, including secure aggregation, differential privacy, encryption-based communication, and defense-aware training strategies. Some methods used in works on glucose-related FL for privacy are those

of Piao *et al.*⁸⁰ where privacy is preserved by exchanging only model parameters and keeping all sensitive data local, and secure aggregation prevents the exposure of individual updates; also, Yang *et al.*⁶⁵, who proposed a personalized FL framework in which models are trained directly on IoT glucose devices. Patient privacy is maintained by local training and selective aggregation, ensuring that personal glucose patterns never leave the device. Finally, Yang *et al.*⁶⁶ improved privacy by aggregating model updates only within patient clusters rather than across all users. By limiting cross-group information flow, the approach reduces privacy leakage while maintaining accurate glucose prediction.

4.5. Federated data analysis

Federated data analysis (FDA) refers to a set of techniques that enable distributed statistical analysis, dataset inspection, and data quality assessment without exposing raw data. In an FDA setting, each institution computes protected local statistics such as averages, variances, feature distributions, or missing-data patterns, which are then securely aggregated to obtain global insights. FDA is particularly important in healthcare applications, where datasets must be validated before being used for learning.

Several studies emphasize the role of FDA in preparing decentralized datasets for reliable FL. Recent work has expanded the scope of FDA by Elkordy *et al.*⁹⁸ who have described secure aggregation protocols that support cross-silo statistical queries while maintaining privacy guarantees.⁵⁹ Wang *et al.*⁹⁹ have introduced a different private federated density estimation method that enables flexible analysis under bounded privacy budgets. Taken together, these studies position FDA as a core component of modern FL pipelines, supporting data consistency, quality assurance, and confidentiality while enabling scalable distributed learning.

4.6. Limitations of federated learning

Despite its advantages, FL presents several limitations that affect the feasibility of deployment and model performance in practice. One of the most persistent challenges is data

Table 4. Privacy protection strategies in federated learning and their relevance to glucose and diabetes applications

FL Paper	Protection used	Relevance to glucose/diabetes FL
Thakur <i>et al.</i> ⁹⁷	Knowledge abstraction and filtering	Keeps detailed glucose features local while still allowing learning across sites
Elkordy <i>et al.</i> ⁹⁸	Secure aggregation for federated analytics	Allows hospitals to work together on glucose analysis without sharing patient data
Wang <i>et al.</i> ⁹⁹	Differential privacy	Adds noise to protect individual glucose patterns during federated analysis
Fung <i>et al.</i> ⁹³	Robust aggregation against poisoning.	Limits the effect of faulty or malicious glucose devices on model training
Kairouz <i>et al.</i> ⁶³	Secure aggregation + differential privacy	Shows how privacy tools can be combined in real healthcare federated systems

Abbreviation: FL: Federated learning.

heterogeneity, where each client holds data drawn from substantially different distributions. Such non-IID settings often lead to biased global models, slower convergence, and unstable training behavior, which is particularly problematic in healthcare applications where reliability is critical.¹⁰⁰

Another major concern is communication overhead. Federated training requires repeated transmission of model parameters or updates between clients and the server, which can place a heavy burden on network bandwidth and require consistent connectivity. In resource-constrained environments or on low-power medical devices, this communication cost can quickly become a bottleneck.¹⁰¹

Although raw data remains local, FL is not immune to privacy and security threats. Shared gradients and parameters may still leak sensitive information or be exploited through adversarial manipulation, including data- and model-poisoning attacks. In addition, real-world systems must contend with straggling clients, intermittent participation, client dropouts, and stale updates, all of which complicate synchronization and slow learning progress.^{61,102} These practical limitations have motivated continued research into personalized, communication-efficient, and security-aware FL methods.

Dataset availability presents an additional barrier to reproducibility in FL for glucose applications. The OhioT1DM dataset, covering 12 participants across the 2018 and 2020 cohorts, remains one of the few openly accessible CGM benchmarks, and only a small number of the reviewed FL studies make use of it. The majority rely on synthetic data, proprietary records, or restricted databases such as MIMIC-III that require formal access agreements. Offline RL trajectory datasets used alongside FL pipelines are similarly unavailable for public use. Without shared, standardized benchmarks, independently verifying reported results or assessing model robustness across different patient populations and device ecosystems remains difficult.

Overall, FL provides a practical framework for developing glucose prediction and diabetes-management models in privacy-sensitive, distributed environments where centralized data sharing is restricted. Beyond the standard client-server design, architectural variants such as hierarchical, personalized, split, graph-based, asynchronous, and decentralized FL offer different trade-offs between scalability, personalization, communication cost, and privacy protection. However, real-world adoption in healthcare remains limited by non-IID data heterogeneity, communication overhead, and persistent security risks such as poisoning and inference attacks. These challenges motivate continued work on robust

aggregation, privacy-preserving analytics, and personalized FL strategies, particularly for patient-specific glucose dynamics and closed-loop decision-support systems.

5. Federated reinforcement learning for glucose control

Federated reinforcement learning (FRL) extends federated learning to reinforcement learning settings, in which decisions are sequential, and outcomes depend on long-term interactions with the environment. Unlike supervised federated learning, FRL focuses on learning control policies rather than static prediction models. This distinction is particularly important in healthcare and especially in glucose management because insulin dosing decisions affect future glucose trajectories through delayed pharmacokinetics and pharmacodynamics, and because the control objective must balance glycemic performance, institutional constraints, and patient-level data sensitivity.^{103,104} Figure 9 shows the FRL architecture for collaborative closed-loop insulin dosing across distributed clinical controllers.

As the trend toward increasingly interconnected devices continues to grow, Sankaradass *et al.*¹⁰⁵ have introduced a novel FL–DRL system for healthcare that integrates both FL and deep RL to enable scalable, privacy-preserving, and real-time analytics. Federated learning ensures that data remain localized and secure, while RL supports real-time decision-making for clinical tasks such as diagnosis and treatment planning.

The authors evaluate their framework using the MIMIC-III dataset, which is preprocessed through several steps, including data cleaning, feature extraction, and normalization. In conclusion, extensive statistical analysis demonstrates that the proposed model outperforms standard baseline models, achieving approximately a 10% improvement in prediction accuracy and a 15% reduction in computational time. These results indicate that the proposed framework effectively addresses key challenges in intelligent healthcare analytics, including data privacy, scalability, and real-time clinical decision support.

Seid *et al.*¹⁰⁶ have proposed the use of unmanned aerial vehicles (UAVs) in scenarios where terrestrial mobile edge computing (MEC) servers are overloaded or unavailable during emergencies. In this work, a multi-UAV-enabled healthcare system is designed using a multi-agent federated reinforcement learning (MAFRL)–based resource allocation framework.

The computation offloading and resource allocation problems are formulated as a Markov Decision Process game within an FL setup involving multiple agents or

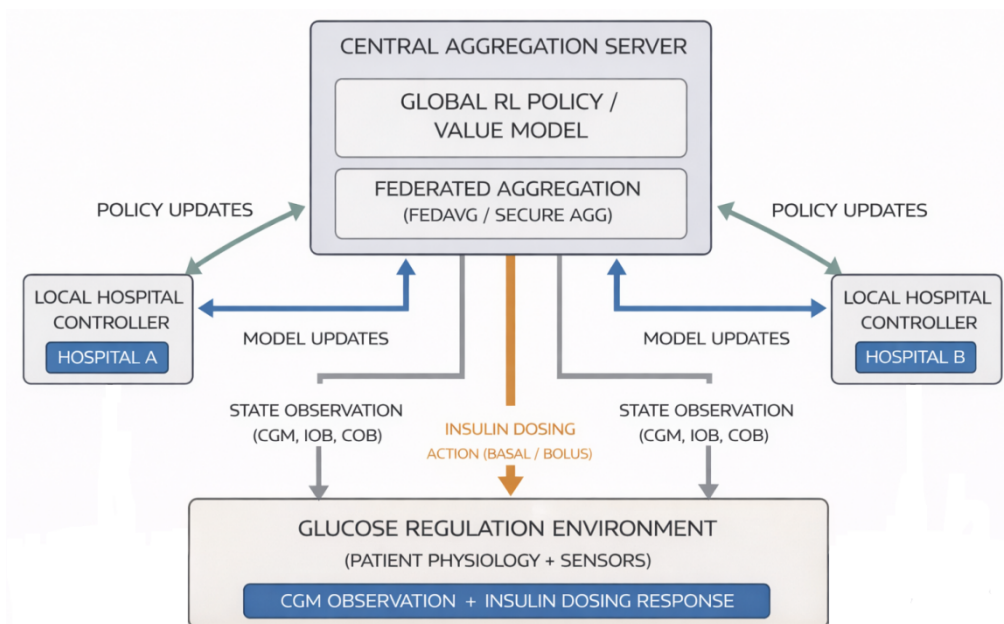


Figure 9. Federated reinforcement learning architecture for collaborative closed-loop insulin dosing across distributed clinical controllers. Image created by the authors.

Abbreviations: COB: Carbohydrates-on-board; CGM: Continuous glucose monitoring; IOB: Insulin-on-board; RL: Reinforcement learning.

participants. The authors then propose an MAFRL algorithm to solve the formulated problem by deciding which tasks to offload, allocating computing resources, and reducing both latency and energy consumption. Extensive simulation results on a real-world heartbeat dataset demonstrate that the proposed system significantly reduces operational costs, preserves data privacy, and improves accuracy relative to baseline learning algorithms.

In another study by Xue *et al.*,¹⁰⁷ the authors have proposed an IoT-enabled e-health system that supports personalized clinical treatment through in-home smart devices for continuous health monitoring. The collected personal health information is used by clinical decision support systems (CDSSs) to assist clinicians in treatment planning. To address issues arising from network congestion and low responsiveness during implementation, the authors integrate mobile edge computing (MEC) and software-defined networking to better leverage the computational and storage capabilities of edge nodes. Based on this architecture, each node deploys a double deep Q-network (DDQN) to learn stable, sequential clinical treatment policies, and a fully decentralized FL framework aggregates DDQN models and extracts shared knowledge across all edge nodes. To address privacy concerns, two additively homomorphic encryption schemes are employed to prevent leakage of electronic medical record data during training. Experimental results show that the proposed model achieves promising performance in

clinical treatment decision support.

In addition, Ahmed *et al.*¹⁰⁸ introduced a work on urinary tract infection detection using an FLRRL model, which can be extended to glucose systems. The authors addressed privacy concerns and the time-intensive nature of testing large datasets by using a combinatorial optimization approach, in which accuracy is optimized as a concave function and computational cost is minimized as a convex function. They employ a federated learning and reinforcement learning strategy (FLRLS framework), leveraging laboratory urine data. The model uses deterministic agents to optimize and exploit the data, while the final decision-making is performed at a centralized aggregation node. Simulations were performed against two baseline algorithms, FLP (Federated Learning Policy) and RFLS (Reinforced Federated Learning Scheme). FLRLS outperformed both baseline models by employing an adaptive, episodic DQN algorithm combined with an FL framework, which allows local states to remain flexible and quickly adapt to changes in resources. Overall, the model demonstrated a notable 5% improvement in accuracy while reducing the overall computational delay.

Also, Ahmed *et al.*¹⁰⁹ have introduced a diabetes prediction technique using an adaptive FL framework to overcome constraints in decentralized settings. They proposed integrating a reinforcement learning framework with tabular Q-learning and DQN strategies for client selection, dynamically identifying clients

that most positively affect global model performance. The RL framework consists of a novel multi-objective reward function that guides the RL agent toward performance- and efficiency-aware client selection. For local training, Random Forest classifiers are employed to ensure robustness to noise, class imbalance, and limited computational resources, particularly in privacy-sensitive healthcare settings. The model is trained on two real-world healthcare datasets, BRFSS2015 and Diabetes Prediction, and is further extended to benchmark federated learning datasets (CIFAR-10 and FEMNIST) to assess scalability and generalization, and compared against baseline models such as FedAvg and tabular Q-learning-based selection.

The proposed adaptive FL model achieves the best overall performance, reaching a global accuracy of 91.3% compared to tabular Q-learning, and demonstrates improvements across both healthcare datasets. It also achieves a 15% reduction in execution time relative to baseline models while maintaining low execution variance and communication overhead. These results demonstrate the effectiveness of the proposed model in balancing predictive performance with system-level efficiency, establishing its suitability for deployment in practical, privacy-sensitive, and resource-constrained healthcare

environments.

Finally, for an FRL architecture, Otoum *et al.*¹¹⁰ described a rigorous intrusion detection system (IDS) in which an FRL-based IDS (FRL-IDS) is introduced within an IoT framework to address the security risks inherent in such systems. The system trains intrusion-detection nodes locally in a distributed environment using a Q-learning-based policy for decision-making, and aggregates model updates from each node via FL with FedAvg. The model is extensively trained on the CICIDS2017 intrusion detection dataset, which is widely used for IDS evaluation. Final results indicate that the proposed model outperforms baseline models such as Support Vector Machine, achieving higher accuracy and detection rates. This system not only supports intrusion detection in large healthcare networks but is also applicable to other wireless decentralized applications, including glucose-related healthcare systems. Overall, the diverse applications of federated reinforcement learning and hybrid RL–FL frameworks in healthcare demonstrate their effectiveness in addressing privacy preservation, resource efficiency, personalized decision making, and system robustness. A comparative overview of these approaches, including their RL techniques, federated architectures, datasets, and relevance to glucose management, is presented in [Table 5](#).

Table 5. Summary of federated reinforcement learning and RL + FL approaches in healthcare

Application domain	RL type	FL architecture	Dataset	Key objective	Relevance to glucose management
IoT e-health ¹¹⁰	DDQN	Decentralized FL	EMR data ¹¹⁰	Personalized treatment policies	Illustrates how home devices can learn personalized treatment strategies, relevant for daily glucose management
Emergency healthcare ¹¹¹	Multi-agent RL	MAFRL	Heartbeat dataset ¹¹¹	Resource allocation & offloading	Demonstrates how multiple agents can coordinate in real time, similar to coordinating sensors and insulin delivery devices
Clinical analytics ¹¹²	Deep DRL (FRL)	Centralized FL	MIMIC-III ¹¹²	Privacy-preserving decision support	Shows how models can learn treatment decisions without sharing patient data, which is useful for insulin dosing systems
UTI detection ¹⁰⁸	DQN	Centralized FL	Urine lab data ¹⁰⁸	Accuracy–cost optimization	Shows how lab-based decision models can be adapted to glucose test data to improve prediction accuracy
Diabetes prediction ¹¹⁴	Q-learning + DQN	Adaptive FL	BRFSS ¹¹³	Diabetes prediction, client selection & efficiency	Directly relevant to glucose systems by selecting the most useful patient data to improve overall model performance
Intrusion detection	Q-learning	Centralized FRL	CICIDS ¹¹⁴	Secure anomaly detection	Highlights how learning systems can be protected from attacks, which is important for safe glucose control devices

Abbreviations: DDQN: Double deep Q-network; DQN: Deep Q-network; EMR: Electronic medical record; FL: Federated learning; FRL: Federated reinforcement learning; IoT: Internet of Things; MAFRL: Multi-agent federated reinforcement learning; RL: Reinforcement learning; UTI: Urinary tract infection.

6. Conclusion

Reinforcement learning approaches for glucose control have matured considerably in simulation, yet real-world validation remains limited. Nearly all controllers spanning policy-gradient methods such as PPO, value-based approaches including DQN, offline RL, and modular architectures are trained and evaluated on virtual cohorts of 20–30 patients using the UVA/Padova or Simglucose simulators, with only a small number validated on real clinical data. Among offline RL methods, TD3-BC demonstrated the strongest real-world performance, improving TIR from 69.89% to 74.38% and reducing time-below-range from 3.59% to 2.73% on a 100-patient real-world DBLG1 dataset. Carbohydrate-agnostic and modular controllers report a broad performance range *in silico*: G2P2C achieved adult TIR in the low-to-mid 70% and adolescent TIR in the low-to-mid 60% range under a demanding 180 g/day carbohydrate protocol, while the bioinspired controller reported approximately 90% TIR with mean glucose near 125 mg/dL in single-meal evaluations. Multi-agent basal-bolus advisors achieved 78–94% TIR alongside a reduction in hypoglycemia from approximately 14–17% to 2–4%. While these results are encouraging, they must be interpreted in the context of the specific simulator setup, cohort composition, and evaluation protocol used in each study, as these factors vary substantially across the literature and limit direct cross-study comparison.

Federated learning offers a practical framework for privacy-preserving collaboration across patients, devices, and institutions, yet persistent challenges remain. Personalized and clustering-based architectures show meaningful improvements under heterogeneous patient data: clustering-based federated deep learning achieved an RMSE of 11.08 ± 1.77 mg/dL for glucose prediction, while FL-aided CNN anomaly detection reached a recall of 99.24% in a diabetes management control system. Privacy-preserving aggregation methods such as TAPFed reduced transmission overhead by 29–45% relative to standard secure aggregation. Despite these advances, non-IID data heterogeneity, communication overhead, and vulnerability to poisoning and inference attacks remain persistent barriers to clinical deployment, and federated learning alone is not sufficient for safety-critical decision-making without careful system-level design and evaluation.

The integration of RL and FL into unified frameworks remains an emerging direction, particularly for glucose-specific applications. From a system perspective, combining reinforcement learning with federated learning may provide a practical pathway toward more robust learning-based glucose control. Reinforcement learning enables

adaptive insulin dosing policies that respond to patient variability and dynamic glucose changes, while federated learning allows models to be trained collaboratively across distributed devices or clinical sites without sharing raw patient data. By leveraging distributed training across heterogeneous patient populations, such architectures may reduce overfitting to simulator environments, support better personalization, and enable privacy-preserving model development across healthcare devices and institutions. Early results in adjacent healthcare domains are encouraging an FL–DRL system evaluated on MIMIC-III achieved approximately a 10% improvement in prediction accuracy alongside a 15% reduction in computational time relative to standard baselines yet glucose-specific federated reinforcement learning remains underexplored, and only a small subset of the reviewed studies directly combine adaptive insulin dosing with privacy-preserving distributed training.

Taken together, these findings indicate that RL-based glucose control requires broader real-world validation beyond simulation, standardized safety benchmarks, and more consistent reporting of evaluation protocols to support clinical translation. The convergence of RL and FL represents a compelling direction for scalable, personalized, and privacy-aware closed-loop systems one that the field is only beginning to explore systematically. Future work should prioritize real patient validation, sim-to-real transfer strategies, and unified RL–FL frameworks designed around the safety and regulatory requirements of next-generation AID systems.

Acknowledgments

None.

Funding

This work was supported in part through Caterpillar Fellowship from Bradley University

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Sherif Abdelfattah, Babu Baniya

Visualization: Sherif Abdelfattah, Babu Baniya

Writing–original draft: Rahul Biswa Karma, Rohit Biswa Karma

Writing–review & editing: Sherif Abdelfattah

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Alberti KGMM, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO Consultation. *Diabet Med*. 1998;15(7):539-553.
doi: 10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S
2. World Health Organization. Diabetes. Available from <https://www.who.int/health-topics/diabetes> [Last accessed on January 28, 2026].
3. Singh A. Basal-bolus injection regimen. Diabetes. Updated 2023. Available from <https://www.diabetes.co.uk/insulin/basal-bolus.html> [Last accessed on January 28, 2026].
4. Digitale E. New research keeps diabetics safer during sleep. Stanford Medicine. 2014. Available from <https://med.stanford.edu/news/insights/2014/05/new-research-keeps-diabetics-safer-during-sleep.html> [Last accessed on January 28, 2026].
5. The Diabetes Control and Complications Trial Research Group. Resource utilization and costs of care in the diabetes control and complications trial. *Diabetes Care*. 1995;18(11):1468-1478.
doi: 10.2337/diacare.18.11.1468
6. Snell-Bergeon JK, Wadwa RP. Hypoglycemia, diabetes, and cardiovascular disease. *Diabetes Technol Ther*. 2012;14(S1):S51-S58.
doi: 10.1089/dia.2012.0031
7. Sumner CJ, Sheth S, Griffin JW, Cornblath DR, Polydefkis M. The spectrum of neuropathy in diabetes and impaired glucose tolerance. *Neurology*. 2003;60(1):108-111.
doi: 10.1212/WNL.60.1.108
8. Coffen RD, Dahlquist LM. Magnitude of type 1 diabetes self-management in youth: health care needs diabetes educators. *Diabetes Educ*. 2009;35(2):302-308.
doi: 10.1177/0145721708327534
9. Raheb MA, Niazmand VR, Eqra N, Vatankeh R. Subcutaneous insulin administration by deep reinforcement learning for blood glucose level control of type 2 diabetic patients. *Comput Biol Med*. 2022;148:105860.
doi: 10.1016/j.compbiomed.2022.105860
10. Sherwood A. Can you reverse type 2 diabetes? WebMD. Updated 2024. Available from: <https://www.webmd.com/diabetes/can-you-reverse-type-2-diabetes> [Last accessed on January 28, 2026].
11. Piras de Oliveira C, Mitchell BD, Fan L, et al. Patient perspectives on the use of half-unit insulin pens by people with type 1 diabetes: a cross-sectional observational study. *Curr Med Res Opin*. 2021;37(1):45-51.
doi: 10.1080/03007995.2020.1843423
12. Forlenza GP, Lal RA. Current status and emerging options for automated insulin delivery systems. *Diabetes Technol Ther*. 2022;24(5):362-371.
doi: 10.1089/dia.2021.0514
13. Sherr JL, Cengiz E, Palerm CC, et al. Reduced hypoglycemia and increased time in target using closed-loop insulin delivery during nights with or without antecedent afternoon exercise in type 1 diabetes. *Diabetes Care*. 2013;36(10):2909-2914.
doi: 10.2337/dc13-0010
14. Shin J, Badgwell TA, Liu KH, Lee JH. Reinforcement learning – overview of recent progress and implications for process control. *Comput Chem Eng*. 2019;127:282-294.
doi: 10.1016/j.compchemeng.2019.05.029
15. Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. 2nd ed. MIT Press; 2018. Available from <https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf> [Last accessed on January 28, 2026].
16. Fox I, Lee J, Pop-Busui R, Wiens J. Deep reinforcement learning for closed-loop blood glucose control. In: Proceedings of the 5th Machine Learning for Healthcare Conference. 2020;126:508-536. Available from <https://proceedings.mlr.press/v126/fox20a.html> [Last accessed on January 28, 2026].
17. Wilinska ME, Chassin LJ, Acerini CL, Allen JM, Dunger DB, Hovorka R. Simulation environment to evaluate closed-loop insulin delivery systems in type 1 diabetes. *J Diabetes Sci Technol*. 2010;4(1):132-144.
doi: 10.1177/193229681000400117
18. Shan Y, Yu J. Non-invasive blood glucose monitoring via multimodal features fusion with interpretable machine learning. *Appl Sci*. 2026;16(2):790.
doi: 10.3390/app16020790
19. Tassa Y, Doron Y, Muldal A, et al. DeepMind control suite. *arXiv*. Preprint posted online 2018.
doi: 10.48550/arXiv.1801.00690
20. Ahmad S, Beneyto A, Zhu T, Contreras I, Georgiou P, Vehi J. An automatic deep reinforcement learning bolus calculator for automated insulin delivery systems. *Sci Rep*. 2024;14(1):15245.
doi: 10.1038/s41598-024-62912-4

21. Man CD, Micheletto F, Lv D, Breton M, Kovatchev B, Cobelli C. The UVA/PADOVA type 1 diabetes simulator: new features. *J Diabetes Sci Technol*. 2014;8(1):26–34.
doi: 10.1177/1932296813514502
22. De Ferdinando MD, Pepe P, De Gennaro S, Palumbo P. Sampled-data static output feedback control of the glucose-insulin system. *IFAC PapersOnLine*. 2020;53(2):3626–3631.
doi: 10.1016/j.ifacol.2020.12.2044
23. Lehmann ED, Deutsch T. A physiological model of glucose-insulin interaction in type 1 diabetes mellitus. *J Biomed Eng*. 1992;14(3):235–242.
doi: 10.1016/0141-5425(92)90058-S
24. Shifrin M, Siegelmann H. Near-optimal insulin treatment for diabetes patients: a machine learning approach. *Artif Intell Med*. 2020;107:101917.
doi: 10.1016/j.artmed.2020.101917
25. Lee J, Kim S, Park SW, Jin SM, Park SM. Toward a fully automated artificial pancreas system using a bioinspired reinforcement learning design: in silico validation. *IEEE J Biomed Health Inform*. 2021;25(2):536–546.
doi: 10.1109/JBHI.2020.3002022
26. Bequette BW. Glucose clamp algorithms and insulin time-action profiles. *J Diabetes Sci Technol*. 2009;3(5):1005–1013.
doi: 10.1177/193229680900300503
27. Hovorka R, Shojaee-Moradie F, Carroll PV, Chassin LJ, Gowrie IJ, Jackson NC, et al. Partitioning glucose distribution, transport, disposal, and endogenous production during IVGTT. *Am J Physiol Endocrinol Metab*. 2002;282(5):E992–E1007.
doi: 10.1152/ajpendo.00304.2001
28. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. *arXiv*. Preprint posted online 2017.
doi: 10.48550/arXiv.1707.06347
29. Marchetti A, Sasso D, D'Antoni F, et al. Deep reinforcement learning for type 1 diabetes: dual PPO controller for personalized insulin management. *Comput Biol Med*. 2025;191:110147.
doi: 10.1016/j.compbiomed.2025.110147
30. Zhao YF, Chaw JK, Ang MC, et al. A safe-enhanced fully closed-loop artificial pancreas controller based on deep reinforcement learning. *PLoS ONE*. 2025;20(1):e0317662.
doi: 10.1371/journal.pone.0317662
31. Lange S, Gabel T, Riedmiller M. Batch reinforcement learning. In: Wiering, M., van Otterlo, M. (eds) *Reinforcement Learning. Adaptation, Learning, and Optimization*. Springer; 2012;12:45–73.
doi: 10.1007/978-3-642-27645-3_2
32. Levine S, Kumar A, Tucker G, Fu J. Offline reinforcement learning: tutorial, review, and perspectives on open problems. *arXiv*. Preprint posted online 2020.
doi: 10.48550/arXiv.2005.01643
33. Emerson H, Guy M, McConville R. Offline reinforcement learning for safer blood glucose control in people with type 1 diabetes. *J Biomed Inform*. 2023;142:104376.
doi: 10.1016/j.jbi.2023.104376
34. Fujimoto S, Meger D, Precup D. Off-policy deep reinforcement learning without exploration. In: Proceedings of the 36th International Conference on Machine Learning; 2019;97:2052–2062. Available from: <https://proceedings.mlr.press/v97/fujimoto19a.html> [Last accessed on January 28, 2026].
35. Kumar A, Zhou A, Tucker G, Levine S. Conservative Q-Learning for offline reinforcement learning. *arXiv*. Preprint posted online 2020.
doi: 10.48550/arXiv.2006.04779
36. Fujimoto S, Gu SS. A minimalist approach to offline reinforcement learning. *arXiv*. Preprint posted online 2021.
doi: 10.48550/arXiv.2106.06860
37. Emerson H, James SG, Guy M, McConville R. Flexible blood glucose control: Offline reinforcement learning from human feedback. *arXiv*. Preprint posted online 2025.
doi: 10.48550/arXiv.2501.15972
38. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature*. 2015;518(7540):529–533.
doi: 10.1038/nature14236
39. Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double Q-learning. *Proc Conf AAAI Artif Intell*. 2016;30(1):2094–2100.
doi: 10.1609/aaai.v30i1.10295
40. Wang Z, Schaul T, Hessel M, Hasselt H, Lanctot M, Freitas N. Dueling network architectures for deep reinforcement learning. In: Proceedings of The 33rd International Conference on Machine Learning. 2016;48:1995–2003. Available from: <https://proceedings.mlr.press/v48/wangf16.html> [Last accessed on January 28, 2026].
41. Schaul T, Quan J, Antonoglou I, Silver D. Prioritized experience replay. *arXiv*. Preprint posted online 2016.
doi: 10.48550/arXiv.1511.05952
42. Hessel M, Modayil J, Van Hasselt H, Schaul T, Ostrovski G, Dabney W, et al. Rainbow: combining improvements in deep reinforcement learning. *Proc Conf AAAI Artif Intell*. 2018;32(1):3215–3222.
doi: 10.1609/aaai.v32i1.11796
43. Bellemare MG, Dabney W, Munos R. A distributional

- perspective on reinforcement learning. In: Proceedings of the 34th International Conference on Machine Learning; 2017:70:449–458. Available from: <https://proceedings.mlr.press/v70/bellemare17a.html> [Last accessed on January 28, 2026].
44. Wang S, Gu W. An improved strategy for blood glucose control using multi-step deep reinforcement learning. In: Proceedings of the 2024 16th International Conference on Bioinformatics and Biomedical Technology. ACM; 2024:196–203.
doi: 10.1145/3674658.3674689
45. Symeonidis P, Rizos E, Andras C, Hairistanidis S, Manolopoulos Y, Zanker M. Deep reinforcement learning for personalized insulin dosing and glucose control of hospitalized ICU patients. *Int J Data Sci Anal.* 2025;20(7):6841–6854.
doi: 10.1007/s41060-025-00857-1
46. Beolet T, Adenis A, Hunecker E, Louis M. End-to-end offline reinforcement learning for glycemia control. *Artif Intell Med.* 2024;154:102920.
doi: 10.1016/j.artmed.2024.102920
47. Hettiarachchi C, Malagutti N, Nolan CJ, Suominen H, Daskalaki E. G2P2C: a modular reinforcement learning algorithm for glucose control by glucose prediction and planning in type 1 diabetes. *Biomed Signal Process Control.* 2024;90:105839.
doi: 10.1016/j.bspc.2023.105839
48. Adjevi A, Abdirashid AM, Aktaş F, Ucar MHB, Solak S. Explainable reinforcement learning for glucose monitoring based on Shapley value analysis. *Comput Methods Programs Biomed.* 2026;278:109266.
doi: 10.1016/j.cmpb.2026.109266
49. Jaloli M, Cescon M. Basal–bolus advisor for type 1 diabetes patients using multi-agent reinforcement learning methodology. *Control Eng Pract.* 2024;142:105762.
doi: 10.1016/j.conengprac.2023.105762
50. Dénes-Fazakas L, Szilágyi L, Kovács L, De Gaetano A, Eigner G. Reinforcement learning: a paradigm shift in personalized blood glucose management for diabetes. *Biomedicines.* 2024;12(9):2143.
doi: 10.3390/biomedicines12092143
51. Dulac-Arnold G, Levine N, Mankowitz DJ, et al. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. *Mach Learn.* 2021;110(9):2419–2468.
doi: 10.1007/s10994-021-05961-4
52. Packer C, Gao K, Kos J, Krähenbühl P, Koltun V, Song D. Assessing generalization in deep reinforcement learning. *arXiv.* Preprint posted online 2018.
doi: 10.48550/arXiv.1810.12282
53. Kovatchev BP, Breton M, Man CD, Cobelli C. *in silico* preclinical trials: a proof of concept in closed-loop control of type 1 diabetes. *Diabetes Sci Technol.* 2009;3(1):44–55.
doi: 10.1177/193229680900300106
54. Yu X, Yang Z, Sun X, Liu H, Li H, Lu J et al. Deep reinforcement learning for automated insulin delivery systems: algorithms, applications, and prospects. *AI.* 2025;6(5):87.
doi: 10.3390/ai6050087
55. McMahan HB, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. *arXiv.* Preprint posted online 2016.
doi: 10.48550/arXiv.1602.05629
56. Nguyen DC, Ding M, Pathirana PN, Seneviratne A, Li J, Niyato D, et al. Federated learning for industrial internet of things in future industries. *IEEE Wirel Commun.* 2021;28(6):192–199.
doi: 10.1109/MWC.001.2100102
57. Yang S, Varghese P, Stephenson E, Tu K, Gronsbell J. Machine learning approaches for electronic health records phenotyping: a methodical review. *J Am Med Inform Assoc.* 2022;30(2):367–381.
doi: 10.1093/jamia/ocac216
58. Guo W, Zhuang F, Zhang X, Tong Y, Dong J. A comprehensive survey of federated transfer learning: challenges, methods and applications. *Front Comput Sci.* 2024;18(6):186356.
doi: 10.1007/s11704-024-40065-x
59. Moore W, Frye S. Review of HIPAA, part 1: history, protected health information, and privacy and security rules. *J Nucl Med Technol.* 2019;47(4):269–272.
doi: 10.2967/jnmt.119.227819
60. Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ Digit Med.* 2020;3(1):119.
doi: 10.1038/s41746-020-00323-1
61. Zhu H, Xu J, Liu S, Jin Y. Federated learning on non-IID data: A survey. *Neurocomputing.* 2021;465:371–390.
doi: 10.1016/j.neucom.2021.07.098
62. Ye M, Fang X, Du B, Yuen PC, Tao D. Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Comput Surv.* 2024;56(3):1–44.
doi: 10.1145/3625558
63. Kairouz P, McMahan HB, Avenet B, et al. Advances and open problems in federated learning. *Found Trends Mach Learn.* 2021;14(1–2):1–210.
doi: 10.1561/22000000083
64. Konečný J, McMahan HB, Ramage D, Richtárik P. Federated optimization: Distributed machine learning for on-device

- intelligence. *arXiv*. Preprint posted online 2016.
doi: 10.48550/arXiv.1610.02527
65. Astillo PV, Duguma DG, Park H, Kim J, Kim B, You I. Federated intelligence of anomaly detection agent in IoTMD-enabled diabetes management control system. *Future Gener Comput Syst*. 2021;128:395–405.
doi: 10.1016/j.future.2021.10.023
66. Jaramillo-Velez D, Rajput C, Freij-Hollanti R, Hollanti C, Amat AG. Perfectly-private analog secure aggregation in federated learning. In: Proceedings of the 2025 IEEE Information Theory Workshop (ITW). IEEE; 2025:1–6.
doi: 10.1109/ITW62417.2025.11240508
67. Xu R, Li B, Li C, Joshi JBD, Ma S, Li J. TAPFed: threshold secure aggregation for privacy-preserving federated learning. *IEEE Trans Dependable Secur Comput*. 2024;21(5):4309–4323.
doi: 10.1109/TDSC.2024.3350206
68. Liu Y, Kang Y, Zhang X, *et al*. A communication-efficient vertical federated learning framework. *arXiv*. Preprint posted online 2019.
doi: 10.48550/arXiv.1912.11187
69. Briggs C, Fan Z, Andras P. Federated learning with hierarchical clustering. In: Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN); IEEE; 2020:1–9.
doi: 10.1109/IJCNN48605.2020.9207469
70. Singh JP, Aqsa A, Ghani I, Sonani R, Govindarajan V. Privacy-aware hierarchical federated learning in healthcare: integrating differential privacy and secure multi-party computation. *Future Internet*. 2025;17(8):345.
doi: 10.3390/fi17080345
71. Arivazhagan MG, Aggarwal V, Singh AK, Choudhary S. Federated learning with personalization layers. *arXiv*. Preprint posted online 2019.
doi: 10.48550/arXiv.1912.00818
72. Yang X, Li J. Edge AI empowered personalized privacy-preserving glucose prediction with federated deep learning. In: Proceedings of the 2023 IEEE International Conference on E-health Networking, Application & Services (Healthcom); IEEE; 2023:224–230.
doi: 10.1109/Healthcom56612.2023.10472368
73. Yang X, Li J. A clustering-based federated deep learning approach for enhancing diabetes management with privacy-preserving edge artificial intelligence. *Healthc Anal*. 2025;7:100392.
doi: 10.1016/j.health.2025.100392
74. Zhang X, Li C, Han C, Li S, Feng Y, Wang H, *et al*. A personalized federated meta-learning method for intelligent and privacy-preserving fault diagnosis. *Adv Eng Inform*. 2024;62:102781.
doi: 10.1016/j.aei.2024.102781
75. Oh J, Kim S, Yun SY. FedBABU: Towards enhanced representation for federated image classification. *arXiv*. Preprint posted online 2022.
doi: 10.48550/arXiv.2106.06042.
76. Dinh C, Tran T, Nguyen TD. Personalized federated learning with Moreau envelopes. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS); 2020;33:21394–21405. Available from: https://proceedings.neurips.cc/paper_files/paper/2020/file/f4f1f13c8289ac1b1ee0ff176b56fc60-Paper.pdf [Last accessed on January 28, 2026].
77. Gupta O, Raskar R. Distributed learning of deep neural network over multiple agents. *J Netw Comput Appl*. 2018;116:1–8.
doi: 10.1016/j.jnca.2018.05.003
78. Alrashed B, Nanda P, Dinh H, Aldahiri A, Alhosaini H, Alghamdi N. PPVFL-SplitNN: privacy-preserving vertical federated learning with split neural networks for distributed patient data. In: Proceedings of the 22nd International Conference on Security and Cryptography; 2025:13–24.
doi: 10.5220/0013445300003979
79. Fu X, Zhang B, Dong Y, Chen C, Li J. Federated graph machine learning: a survey of concepts, techniques, and applications. *ACM SIGKDD Explor Newsl*. 2022;24(2):32–47.
doi: 10.1145/3575637.3575644
80. Piao C, Li K. Blood glucose level prediction: a graph-based explainable method with federated learning. *arXiv*. Preprint posted online 2023.
doi: 10.48550/arXiv.2312.12541
81. Marling C, Bunescu R. The OhioT1DM dataset for blood glucose level prediction. In: Proceedings of the 3rd International Workshop on Knowledge Discovery in Healthcare Data (KDH 2018). CEUR Workshop Proceedings; 2018:60–64. Available from: <https://ceur-ws.org/Vol-2148/paper09.pdf> [Last accessed on January 28, 2026].
82. Marling C, Bunescu R. The OhioT1DM dataset for Blood Glucose Level Prediction: Update 2020. *CEUR Workshop Proc*. 2020;2675:71–74. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC7881904/> [Last accessed on January 28, 2026].
83. Xie C, Koyejo O, Gupta I. Asynchronous federated optimization. *arXiv*. Preprint posted online 2019.
doi: 10.48550/arXiv.1903.03934
84. Piao C, Zhu T, Wang Y, *et al*. Privacy-preserved blood

- glucose level cross-prediction: an asynchronous federated learning approach. *IEEE J Biomed Health Inform.* 2026; 30(2):839–852.
doi: 10.1109/JBHI.2025.3573954
85. Kalra S, Wen J, Cresswell JC, Volkovs M, Tizhoosh HR. Decentralized federated learning through proxy model sharing. *Nat Commun.* 2023;14(1):2899.
doi: 10.1038/s41467-023-38569-4
86. Tripathy SS, Beborra S, Chowdhary CL, Mukherjee T, Kim S, Shafi J, *et al.* FedHealthFog: a federated learning-enabled approach towards healthcare analytics over fog computing platform. *Heliyon.* 2024;10(5):e26416.
doi: 10.1016/j.heliyon.2024.e26416
87. Liang X, Zhao J, Chen Y, Bandara E, Shetty S. Architectural design of a blockchain-enabled, federated learning platform for algorithmic fairness in predictive health care: design science study. *JMIR Med Inform.* 2023;25:e46547.
doi: 10.2196/46547
88. Yuan L, Wang Z, Sun L, Yu PS, Brinton CG. Decentralized federated learning: a survey and perspective. *IEEE Internet Things J.* 2024;11(21):34617–34638.
doi: 10.1109/JIOT.2024.3407584
89. Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. In: Proceedings of Machine Learning and Systems; 2020;2:429–450. Available from: https://proceedings.mlsys.org/paper_files/paper/2020/file/1f5fe83998a09396ebe6477d9475ba0c-Paper.pdf [Last accessed on January 28, 2026].
90. Wang J, Liu Q, Liang H, Joshi G, Poor HV. Tackling the objective inconsistency problem in heterogeneous federated optimization. In: Proceedings of the 33rd Conference on Neural Information Processing Systems; 2020;33:7611–7623. Available from: <https://proceedings.neurips.cc/paper/2020/hash/564127c03caab942e503ee6f810f54fd-Abstract.html> [Last accessed on January 28, 2026].
91. Reddi SJ, Charles Z, Zaheer M, *et al.* Adaptive federated optimization. *arXiv.* Preprint posted online 2021.
doi: 10.48550/arXiv.2003.00295
92. Darpit D, Vyas K, Jayagopal JK, Garcia A, Erraguntla M, Lawley M. FedGlu: A personalized federated learning-based glucose forecasting algorithm for improved performance in glycemic excursion regions FedGlu: Personalized federated-learning based glucose forecasting algorithm. *Res Sq.* Preprint posted online 2025.
doi: 10.21203/rs.3.rs-7339691/v1
93. Fung C, Yoon CJ, Beschastnikh I. The limitations of federated learning in sybil settings. In: Proceedings of the 23rd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2020); 2020:301–316. Available from: <https://www.usenix.org/conference/raid2020/presentation/fung> [Last accessed on January 28, 2026].
94. Zhu L, Liu Z, Han S. Deep leakage from gradients. In: Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019); 2019;32. Available from: <https://proceedings.neurips.cc/paper/2019/hash/60a6c4002cc7b29142def8871531281a-Abstract.html> [Last accessed on January 28, 2026].
95. Shokri R, Stronati M, Song C, Shmatikov V. Membership inference attacks against machine learning models. In: Proceedings of the IEEE Symposium on Security and Privacy; 2017:3–18.
doi: 10.1109/SP.2017.41
96. Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security; 2015:1322–1333.
doi: 10.1145/2810103.2813677
97. Thakur A, Molaei S, Nganjimi PC, *et al.* Knowledge abstraction and filtering based federated learning over heterogeneous data views in healthcare. *NPJ Digit Med.* 2024;7(1):283.
doi: 10.1038/s41746-024-01272-9
98. Elkordy AR, Ezzeldin YH, Han S, *et al.* Federated analytics: A survey. *APSIPA Trans Signal Inf Process.* 2023;12(1):1–33.
doi: 10.1561/116.00000063
99. Wang J, Cao L, Chai C, Li G. Federated data analytics with differentially private density estimation model. In: Proceedings of the 2025 IEEE 41st International Conference on Data Engineering (ICDE); 2025:2768–2781.
doi: 10.1109/ICDE65448.2025.00208
100. Zhao Y, Li M, Lai L, Suda N, Civin D, Chandra V. Federated learning with non-IID data. *arXiv.* Preprint posted online 2018.
doi: 10.48550/arXiv.1806.00582
101. Konečný J, McMahan HB, Yu FX, Richtárik P, Suresh AT, Bacon D. Federated learning: strategies for improving communication efficiency. *arXiv.* Preprint posted online 2016.
doi: 10.48550/arXiv.1610.05492
102. Bagdasaryan E, Veit A, Hua Y, Estrin D, Shmatikov V. How to backdoor federated learning. In: Proceedings of the International Conference on Artificial Intelligence and Statistics; 2020;108:2938–2948. Available from: <https://proceedings.mlr.press/v108/bagdasaryan20a> [Last accessed on January 28, 2026].
103. Shifa, A, Abdelfattah, S, Baniya, B. A Federated Deep Reinforcement Learning Approach for Reliable Glucose

- Regulation. In: Proceedings of IEEE World AI IoT Congress (AIIoT). 2025; 0574-0580.
doi: 10.1109/AIIoT65859.2025.11105348
104. Shifa, A, Abdelfattah, S, Badr, M. M., Baza, M., Rasheed, A. Personalized Federated Deep Reinforcement Learning System for Diabetes Management With LLM Integration. In: Proceedings of 2025 International Telecommunications Conference (ITC-Egypt). 2025; 394-399.
doi: 10.1109/ITC-Egypt66095.2025.11186675
105. Sankaradass V, Manindra Manish VK. Federated learning and deep reinforcement networks for privacy-preserving real-time data analytics in smart healthcare systems. *Cluster Comput.* 2025;28(16):1019.
doi: 10.1007/s10586-025-05738-7
106. Seid AM, Erbad A, Abishu HN, Albaseer A, Abdallah M, Guizani M. Multiagent federated reinforcement learning for resource allocation in UAV-enabled internet of medical things networks. *IEEE Internet Things J.* 2023;10(22):19695–19711.
doi: 10.1109/JIOT.2023.3283353
107. Xue Z, Zhou P, Xu Z, *et al.* A resource-constrained and privacy-preserving edge-computing-enabled clinical decision system: a federated reinforcement learning approach. *IEEE Internet Things J.* 2021;8(11):9122–9138.
doi: 10.1109/JIOT.2021.3057653
108. Ahmed S, Groenli TM, Lakhan A, Chen Y, Liang G. A reinforcement federated learning based strategy for urinary disease dataset processing. *Comput Biol Med.* 2023;163:107210.
doi: 10.1016/j.compbimed.2023.107210
109. Ahmed S, Shamim Kaiser M, Chaki S, Shawkat Ali ABM. Adaptive federated learning with reinforcement learning-based client selection for heterogeneous environments. *IEEE Access.* 2025;13:131671–131695.
doi: 10.1109/ACCESS.2025.3591699
110. Otoum S, Guizani N, Mouftah H. Federated reinforcement learning-supported IDS for IoT-steered healthcare systems. In: Proceedings of the ICC 2021—IEEE International Conference on Communications; 2021:1–6.
doi: 10.1109/ICC42927.2021.9500698
111. Goldberger AL, Amaral LA, Glass L, *et al.* PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation.* 2000;101(23):e215–e220.
doi: 10.1161/01.CIR.101.23.e215
112. Johnson AE, Pollard TJ, Shen L, *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data.* 2016;3(1):160035.
doi: 10.1038/sdata.2016.35
113. Behavioral Risk Factor Surveillance System. CDC–2015 BRFSS Survey Data and Documentation. Centers for Disease Control and Prevention. Published 2015. Available from: https://www.cdc.gov/brfss/annual_data/annual_2015.html [Last accessed on January 29, 2026].
114. Sharafaldin I, Lashkari AH, Ghorbani AA. Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP). 2018:108–116.
doi: 10.5220/0006639801080116