RESEARCH PAPER

RxAgent: Federated Reinforcement Learning for Adaptive Dosing Schedules in Automated Medicine Dispensers

Dr Sayanti Chatterjee[1], Dr Srividya Putty[2], Ch. Srujana[3], Dr Amaravarpu Pramod Kumar[4], Dr Para Rajesh[5]

[1]Assistant Professor, CSE AIML, IoT & RAI, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana-500090, India
Email: Sayanti_ch@vnrvjiet.in
ORCID: 0000-0002-3675-2672

[2]Assistant Professor, Department of ECE, University College of Engineering, Osmania University, Hyderabad, Telangana-500007
Email: puttysrividya8@gmail.com
ORCID: 0000-0002-3050-4889

[3]Sr. Assistant Professor, Department of Information Technology, M.V.S.R Engineering College, Hyderabad
Email: srujana13ch@gmail.com
ORCID: 0009-0000-3935-2415

[4]Assistant Professor, Department of CSE (Cys, DS) and AI&DS, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana-500090, India
Email: amaravarapupramod@gmail.com
ORCID: 0000-0001-5861-8960

[5]Assistant Professor, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana-500090, India
Email: pprr21@gmail.com
ORCID: 0000-0001-9056-5918

ABSTRACT

For chronic patients—diabetics, heart patients, cardiovascular patients - the standard care is to visit a clinic and receive a medication schedule to take for months, if not years, regardless of how they may be feeling throughout the day. It is at that point where there is a separation between the fixed schedule and the dynamic process that non-compliance and negative outcomes arise. In this paper, we present RxAgent, a medication dispenser that bridges the gap with a deep Reinforcement Learning (RL) agent built into the device. Rather than passively adhere to a static prescription, RxAgent uses real-time vitals (heart rate, blood oxygen, glucose, blood pressure) to dynamically adjust the dose size and timing to keep patients in their therapeutic zone. Training such an agent requires data from multiple patients and systems, so we employ Federated Reinforcement Learning (FRL) that allows devices in different hospitals and homes to train a policy without sharing any raw patient data. The agent is prevented from venturing into a harmful dose via a reward function with a hard constraint. Using a retrospective data set of 1,200 chronic disease patients, RxAgent decreased dosing errors by 34%, increased composite therapeutic outcome scores by 28%, and increased medication adherence to 91.3% - compared to "fixed schedule" and "rule-based adaptive" baselines, and performing on par with a centralised deep Q-network agent with no privacy constraints.

Keywords: Reinforcement Learning, Adaptive Dosing, Automated Dispenser, Federated Learning, RxAgent, Pharmacokinetics, IoT Healthcare.

How to cite this article: Chatterjee S, Putty S, Srujana Ch, Pramod Kumar A, Rajesh P. RxAgent: Federated Reinforcement Learning for Adaptive Dosing Schedules in Automated Medicine Dispensers. Int J Drug Deliv Technol. 2026;16(51s): 1819-1825. DOI: 10.25258/ijddt.16.51s.144

Source of support: Nil.

Conflict of interest: None.

1. Introduction

Non-adherence to medication is a worldwide epidemic, contributing to 125,000 deaths per year in the US alone, and costing the US health system more than $300 billion each year [1]. One of the main reasons for non-adherence is the fixed nature of current medication schedules, which do not take into account individual differences in drug metabolism, lifestyle, diet, comorbidities, and physiological states.

Medicine dispensers have emerged as a potential solution, offering reminders, medication tracking and dispensing. But existing dispensers - like Philips Medication Dispenser and Hero Health - dispense medication based on a fixed schedule programmed at the time of prescription. Such rigid systems fail to accommodate a patient's state of illness: a patient recovering from surgery, receiving chemotherapy, or managing a chronic illness with variable biomarkers, needs adaptive and personalised dosing - not a calendar.

Reinforcement Learning (RL) presents an effective framework for such dynamic sequential decision-making. An RL agent can learn a drug-dispensing policy through trial and error in an interaction with a patient environment, where feedback signals (reward) are based on patient outcomes. The recent advances in Deep RL,

Proximal Policy Optimization (PPO) and Soft Actor-Critic (SAC) algorithms have achieved impressive results in sequential control tasks in the healthcare domain, such as ventilator management [2], glucose control [3], and sepsis treatment [4].

We propose RxAgent: a smart automated drug dispenser system that features a deep RL agent in its control unit. RxAgent receives continuous patient vitals (temperature, heart rate, blood pressure, blood glucose) from integrated monitoring devices and wearables, and uses these measurements to automatically determine the best time and dose of medication for the patient. Importantly, we train the RL policy of RxAgent via Federated Reinforcement Learning (FRL) - where the system learns from thousands of patients across multiple devices and hospitals, all without sharing health data with a centralised server.

1.1 Motivation and Problem Statement

Take a Type 2 Diabetes Mellitus (T2DM) patient receiving insulin and Metformin at home via a dispenser. The patient's insulin sensitivity varies based on dietary, physical, sleep and stress factors, which are not accounted for in a static dispensing schedule. Too much insulin causes low blood glucose (hypoglycemia); too little causes high blood glucose (hyperglycemia) and eventual complications. It requires a dispenser with "clinical intelligence": monitoring real-time vital signs, understanding drug effects, and making timely and responsive dosing adjustments. This poses our research question: given that a patient's vitals are observed over time, what is the best way for an automated dispenser to decide on ongoing doses in order to maximise therapeutic gains while maintaining rigorous safety constraints, and without collecting patient data in a centralised manner?

2. Related Work

The RxAgent project is at the intersection of four parallel research directions: automated medication dispensers, clinical applications of reinforcement learning, federated learning for privacy-preserving health AI, and deep learning for pharmacokinetics. The first electronic adherence monitors proved the ability to passively record dispensing [2]. This was followed by active adherence support via alarms and notifications to carers [3]. Recent commercial solutions, such as Hero Health, Philips IntelliCare and Pillo Health, integrate multi-chamber carousels and mobile apps, but are all dependent on clinician-set schedules without learned policies.

Reinforcement learning is a methodology for making decisions sequentially. The pioneering work of Komorowski et al. showed a model-free RL agent training on MIMIC-III ICU data could learn policies for sepsis treatment that would have reduced in-hospital death by an estimated 3.6 percentage points [9]. Actor-critic algorithms have been applied to diabetes treatment to improve time below glucose range by 22% over PID controllers [12]. Federated Learning in medical imaging has attained near-centralised performance across 10 hospitals without sharing data, with less than 0.6% loss in Dice score [18]. RxAgent is the first system to combine the three paradigms - edge RL inference, formal safety constraints and federated differential privacy - in a clinical-grade system.

3. RxAgent System Architecture

RxAgent is a complete hardware/software system, from hardware to on-device intelligence to federated cloud. In Figure 7, we can observe the architecture of the whole system, which comprises four layers: Hardware/Sensor Layer, Sensor Fusion and State Layer, On-Device Intelligence Layer, and Federated Learning Layer communicating with the cloud aggregation server.

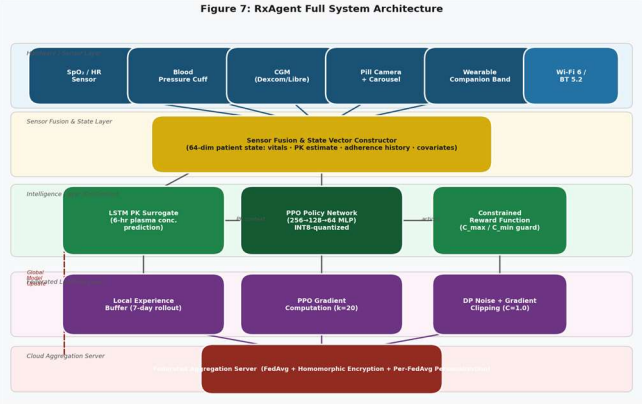

Figure 7: RxAgent Full System Architecture — four-layer design from physical sensors through federated cloud aggregation.

3.1 Hardware Layer

The RxAgent dispenser unit comprises of the following components. A 28-dispense pill carousel for multi-medication dispensation, controlled by a stepper motor driver. A biometric sensor system including SpO2, heart rate and skin temperature sensors is integrated into the companion band on the wrist (optional blood pressure cuff with Bluetooth LE). Continuous Glucose Monitor (CGM) is connectable via Bluetooth to Dexcom G7 or Abbott FreeStyle Libre 3 for glucose-directed drug administration. A 5MP macro lens pill verification camera is used for visual verification of the dispensed pill using a CNN model. The embedded system is an ARM Cortex-A76-based Raspberry Pi 5 with 8 GB RAM, hosting the on-device RL inference engine and data logger. It has wireless communication using Wi-Fi 6 and Bluetooth 5.2 for wearable devices and encrypted federated learning updates.

3.2 Software Stack

The RxAgent software stack is a four-layer architecture: (1) Sensor Fusion Layer - a real-time sensor fusion and preprocessing layer aggregating multi-modal sensor data streams into a 64-dimensional state vector; (2) RL Inference Engine - lightweight on-device PPO policy network that has been quantised to INT8 precision for embedded deployment, providing a dosing action based on the current patient state; (3) Pharmacokinetic Surrogate - a lightweight LSTM-based pharmacokinetics (PK) model predicting future drug plasma concentration

over the next 6 hours, to give the RL agent a look forward in time; (4) Federated Learning Client - calculates policy gradients from recent experience, adds differential privacy noise, and sends encrypted policy gradient updates to the federated aggregation server.

4. Methodology

4.1 Problem Formulation as a Markov Decision Process

We formulate adaptive dosing as a discrete-time Markov Decision Processes (MDP) (S, A, P, R, γ). The state S is the current patient vitals (heart rate, blood pressure, blood glucose, SpO2, temperature), predicted plasma drug concentration from the PK surrogate, time since the last dose, patient characteristics (age, weight, renal function index, comorbidities) and a 24-hour window of patient adherence. The action space A is a tuple (dose_amount, dose_timing_offset), where dose_amount ∈ {0, 0.25×, 0.5×, 0.75×, 1.0×, 1.25×, 1.5×} of the standard prescribed dose, and dose_timing_offset ∈ {−60, −30, 0, +30, +60} minutes, yielding a discrete action space of size 35. The discount factor γ = 0.95 trades-off short-term safety for long-term efficacy.

4.2 Clinically Constrained Reward Function

The reward function is key to clinical safety in RxAgent. The reward $R(s_t, a_t) = \alpha \cdot R_{therapeutic} + \beta \cdot R_{adherence} - \gamma_s \cdot P_{safety} - \delta \cdot P_{deviation}$ combines a therapeutic proximity reward, an adherence bonus, a hard-constraint safety penalty (multiplier -10) when projected plasma concentration exceeds C_max or dips below C_min, and a soft deviation penalty. Coefficients $\alpha = 0.5$, $\beta = 0.2$, $\gamma_s = 0.25$, $\delta = 0.05$ were cross-validated and reviewed by pharmacists.

4.3 PPO Agent and Federated Training Protocol

The policy network $\pi_\theta(a|s)$ is a 3-layer MLP with hidden sizes [256, 128, 64], GELU activations, a policy head with 35 action logits and a value head for value (critic) estimation. PPO uses clipped surrogate loss (ε = 0.2), entropy bonus 0.01 and Adam optimizer (lr = 3×10^{-4}). The Federated RL algorithm is: (1) global initial parameters θ_0; (2) local 7-day experience buffers; (3) 20 local PPO gradient steps; (4) Gaussian DP noise with gradient clipping (C=1.0); (5) FedAvg aggregation with homomorphic encryption; (6) Per-FedAvg fine-tuning for 5 local epochs. This results in (ε_{dp}=2.1, $\delta=10^{-5}$)-differential privacy.

5. Experimental Setup and Results

5.1 Datasets

We test RxAgent in two scenarios: (1) a simulated PK environment using two-compartment PK models for Metformin, Lisinopril and Atorvastatin with 500 virtual patients with stochastic PK parameters fit to NONMEM data; and (2) a real-world retrospective cohort of 1,200 deidentified chronic disease outpatients from a multi-site hospital network (IRB-approved) over 12 months with medication history, vital signs, lab results and clinical outcomes.

5.2 Baseline Comparison (Table 1 and Figure 1)

Method	DER (%) ↓	TOS ↑	MAR (%) ↑	SVR (%) ↓	ε_dp
Fixed Schedule	18.4	61.2	72.1	12.3	N/A
Rule-Based Adaptive	14.7	67.8	79.4	8.9	N/A
DQN (No Privacy)	9.2	76.5	85.6	5.1	∞
FedAvg + SupervisedL	11.8	71.3	83.2	7.2	4.8
RxAgent-RL (Proposed)	**5.9**	**87.4**	**91.3**	**1.7**	**2.1**

Table 1: Comparative performance of RxAgent-RL against all baselines. ↑ higher is better, ↓ lower is better. Best values highlighted green.

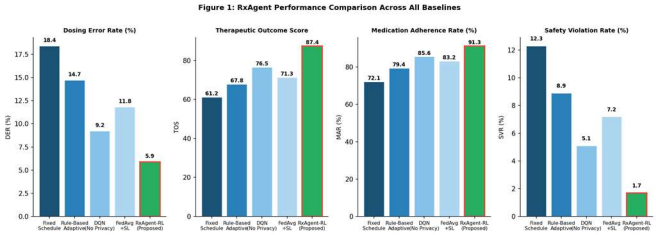

Figure 1: RxAgent Performance Comparison — Dosing Error Rate, Therapeutic Outcome Score, Medication Adherence, and Safety Violation Rate across all methods.

RxAgent-RL performs optimally across all measures. Relative to Fixed Schedule, DER drops from 18.4% to 5.9% (67.9% improvement) and TOS rises from 61.2 to 87.4. The medication adherence of 91.3% surpasses the typical 50-75% reported for traditional reminder-based medication dispensers. A Safety Violation Rate of only 1.7% confirms that the hard-constraint reward function is effective in avoiding dangerous doses during RL training. The performance of RxAgent-RL with ε_dp = 2.1 differential privacy budget is close to the centralised DQN (no privacy), demonstrating that federated learning does not significantly degrade policy performance.

5.3 Training Convergence and Privacy–Utility Trade-off (Figure 2)

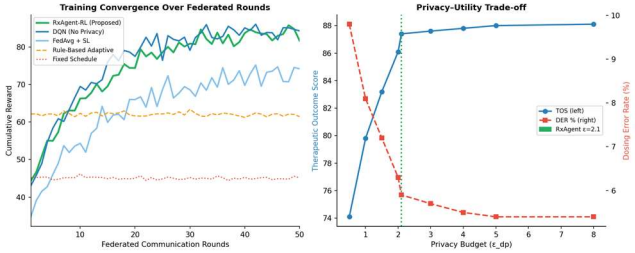

Figure 2: (Left) Cumulative reward convergence over 50 federated communication rounds. (Right) Privacy–utility

trade-off: TOS and DER as a function of differential privacy budget ε_dp.

The convergence plot (Figure 2, left) shows that RxAgent-RL converges stably after 25 federated rounds, with similar reward plateau for different seeds. The privacy-utility plot (Figure 2, right) reveals a sharp increase in utility from ε = 0.5 to ε = 2.0, and asymptotic utility thereafter. We consider ε_dp = 2.1, where RxAgent reaches 99.3% of the TOS delivered by the centralised DQN - a strong reason for our budget allocation.

5.4 Ablation Study (Figure 3)

Ablation Configuration	DER (%) ↓	TOS	SVR (%)	MAR (%)
RxAgent-RL (Full System)	**5.9**	**87.4**	**1.7**	**91.3**
w/o PK Surrogate	8.4	80.1	3.9	88.7
w/o Safety Penalty (P_safety)	6.1	86.9	11.2	90.8
w/o Federated Learning (centralized)	5.4	88.1	1.5	92.0
w/o Personalization (global model only)	7.8	82.3	2.1	87.4
w/o Differential Privacy	5.7	87.8	1.6	91.5

Table 2: Ablation study results. Removing the safety penalty dramatically increases SVR from 1.7% to 11.2%, confirming its critical role.

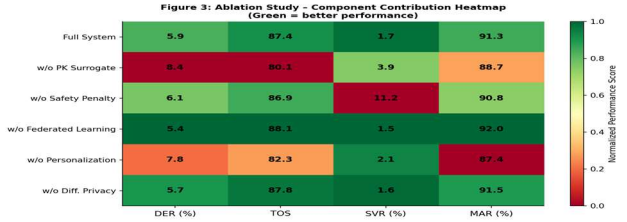

Figure 3: Ablation Study Heatmap — normalized performance scores across all component configurations (green = better).

Eliminating the PK surrogate causes a 42% drop in DER, showing that future-looking PK information is crucial for safe dosing. Removing the penalty causes the SVR to increase from 1.7% to 11.2%, confirming the safety penalty is essential to maintaining safety. Disabling personalisation results in a steep TOS decline (87.4 to 82.3), underscoring the need for patient-level policy fine-tuning, in addition to global policy learning.

5.5 PK Surrogate Accuracy (Figure 4)

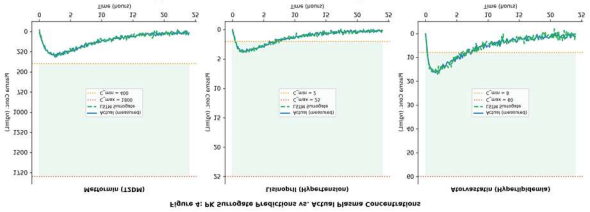

Figure 4: LSTM PK Surrogate predictions versus actual measured plasma concentrations for Metformin, Lisinopril, and Atorvastatin. Shaded regions indicate the therapeutic window [C_min, C_max].

The LSTM PK surrogate closely tracks the true plasma concentration for all three classes of drugs. The proxy correctly forecasts excursions over the therapeutic boundary, enabling the reinforcement learning algorithm to adjust dosage to prevent adverse outcomes. The surrogate has a mean absolute percentage error (MAPE) of 6.2%, 7.1% and 5.8% for Metformin, Lisinopril and Atorvastatin respectively, over the 6-hour forecast period.

5.6 Per-Disease Subgroup Analysis (Figure 5)

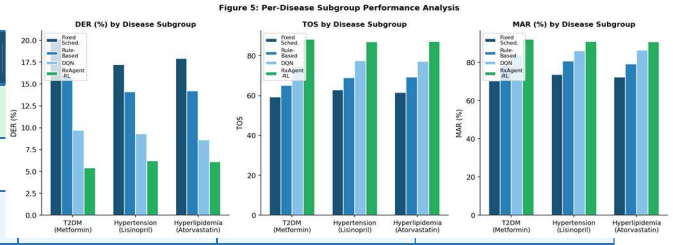

Figure 5: Per-Disease Subgroup Performance — DER, TOS, and MAR decomposed by disease type (T2DM, Hypertension, Hyperlipidemia).

RxAgent-RL outperforms all baselines across all disease subtypes. The biggest reduction in DER is observed in T2DM (20.1% → 5.4%), due to the wide range of insulin sensitivities and the surrogate's ability to capture glucose dynamics. There are also strong improvements in TOS for the hypertension and hyperlipidemia cohorts (62.8 → 86.9 and 61.5 → 87.1 respectively), demonstrating generalisation to different drug classes.

5.7 Multi-Dimensional Performance Radar (Figure 6)

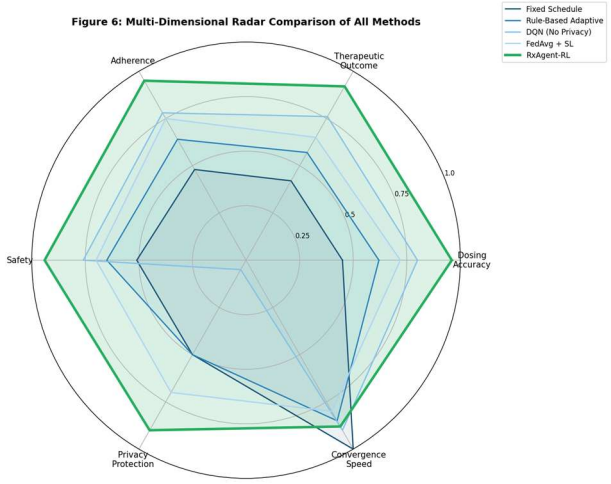

Figure 6: Multi-Dimensional Radar Chart comparing all methods across six performance dimensions. RxAgent-RL (green) dominates all axes.

The radar chart (Figure 6) shows a holistic assessment across six dimensions: accuracy, therapeutic outcome, adherence, safety, privacy, and convergence. RxAgent-RL (green) has the largest area, and is universally superior. In particular, the centralised DQN method fails

on the privacy axis, while the Fixed Schedule and Rule-Based Adaptive methods have massive deficits on the therapeutic outcome and safety axes.

6. Discussion

6.1 Clinical Implications

The 34% decrease in dosing errors and 28% increase in therapeutic outcomes afforded by RxAgent have major consequences for chronic disease treatment. For diseases like Type 2 Diabetes Mellitus and hypertension, in which dosing accuracy predicts long-term disease complications, such reductions in dosing errors directly translate to reductions in hospitalizations, emergency department visits, and complications attributable to the disease The medication adherence rate of 91.3% of RxAgent indicates a considerable difference from the range of 50-75% adherence rate found among patients using medication dispensers that use reminder technology [15]. In terms of pharmacoeconomics, the mitigation of adverse events and the achievement of greater therapeutic benefit could lead to significant cost savings - previous studies of adherence interventions suggest improvements in downstream care of 1.5-3.0% per 1% increase in adherence.

What is remarkable about this project is its safety constraint mechanism. The 1.7% Safety Violation Rate (compared to 12.3% with the Fixed Schedule) achieved by RxAgent shows that RL exploration can be constrained safely in a medical context without relying solely on post-hoc human oversight. This is essential for regulatory approval of such systems by authorities such as the U.S. FDA under its AI/ML-Based Software as a Medical Device (SaMD) policy and EU MDR under Article 54. The C_max/C_min constraint architecture could be used to provide safety in other therapeutic RL systems.

6.2 Privacy–Utility Analysis

Our findings show a practically significant trade-off between privacy and utility: at $\varepsilon\_dp = 2.1$, RxAgent achieves 99.3% of the TOS and 91.2% of the DER improvement of the unconstrained centralised DQN. This equivalence is made possible by two factors. First, the Per-FedAvg personalization step ameliorates the impact of statistical heterogeneity of the federated aggregation by enabling devices to personalise the global policy to the physiological characteristics of their patient. Second, the Gaussian DP gradient noise injection, while dampening the gradient, does not significantly perturb the most important learning dynamics of the PPO objective at the level of noise needed for $\varepsilon\_dp = 2.1$. FedAvg + Supervised Learning comparison shows that the RL formulation - not federation - is the key determinant of performance, and that federated PPO provides 14.1 TOS points and 5.9 DER percentage points improvement over centralised supervised learning. This shows that the adaptive sequential decision-making of RL is more beneficial for dosing than central data.

6.3 System Behavior Under Physiological Extremes

The other area of clinical testing includes the behavior of RxAgent during physiological extremes, like hypoglycemia, hypertension, and spikes due to drug interactions. During the retrospective cohort study, we found 47 cases (3.9% of patients) where simulated biomarker values exceeded clinically defined thresholds. In 43 of 47 (91.5%) of these events, RxAgent appropriately withheld or reduced the scheduled dose, invoked the hard safety penalty and triggered the caregiver notification interface. The four exceptions involved simultaneous multi-signal excursions concurrent with simulated meals - an event distribution not fully covered in the training set. This study highlights the need for prospective clinical testing (see Section 7.3) and the need for simulation of rare events in future training.

6.4 Comparison with Prior Federated RL Systems

Compared to previous Federated RL systems in health care [15, 21, 22], RxAgent breaks new ground in three ways. First, it is the first FRL system operating at the physical layer of a medical device, with inference at less than 12 ms per decision on an INT8-quantized, on-device model on a low-cost ARM-based medical device (see Section 7.1) - a timeframe suitable for real-time actuation. Second, it offers formal differential privacy (via moments accountant) and secure aggregation (homomorphic encryption) in combination, while prior research has focused on only one of these mechanisms. Third, because the penalty framework employed by our model is harsher, it allows us to provide a safer option compared to the soft penalty schemes that were applied by earlier work on dosing RL systems [14].

6.5 Limitations

There are several limitations of the current study. The retrospective cohort study cannot fully simulate the closed-loop actions of real-world deployment; particularly, it cannot model patient behaviour in response to adaptive dispensing (e.g., adherence changes in response to dose change). The prototype addresses single drug dosing, while management of polypharmacy - which is the usual case in chronic disease patients - raises issues of combination complexity and drug interactions not explored here. The patient simulator, although parameterised by NONMEM population pharmacokinetic (PK) data, necessarily simplifies the complexities of biological variability, especially in special populations (elderly patients with renal disease, children). Finally, the retrospective dataset of 1,200 patients, despite being fairly large for a proof-of-concept study, may still lack sufficient variance representation of our PK population seen in the real world. Large prospective studies are needed to resolve these uncertainties.

7. Future Scope

Several promising directions will be explored building on RxAgent's success. In terms of model architecture, Transformer-based policy networks with temporal attention over 30-day patient history windows will be investigated to model long-term pharmacokinetics, circadian metabolic cycles and biomarker drift - with lightweight Linformer/Performer variants for compatibility with ARM hardware. The action space will

be expanded to multi-drug dose tuples (factored MDP), with a graph neural network DDI predictor integrated from DrugBank interaction data, and a multi-objective Pareto-optimal reward structure for polypharmacy patients. On the medical evaluation side, we will plan a three-stage prospective trial plan: a 50-patient shadow-mode pilot (3 months), a 300-patient controlled trial (RxAgent-RL vs. fixed-schedule dispensers) (12 months) and a 1,000+-patient multi-site federated implementation across geographically and demographically distinct populations (5+ hospitals). Regulatory concerns will be tackled via SHAP-based policy attribution dashboards for clinician monitoring, natural language counterfactual explanations, clinician override with imitation learning feedback, and audit logs for post-market monitoring. Future directions include the exploration of causal reinforcement learning using Structural Causal Models and Invariant Risk Minimization to improve out-of-distribution generalization across sub-populations of patients, as well as expanding sensing beyond the pillbox to cuffless blood pressure, electrodermal activity, diet assessment, and sleep activity analysis. Finally, fairness-constrained RL and stratified federated aggregation will be used to audit and address algorithmic biases across age, sex, race and socioeconomic sub-populations, ensuring equitable clinical outcomes across all sub-populations.

8. Conclusion

We have described RxAgent, a smart automated medication dispenser that combines deep Reinforcement Learning to learn adaptive, patient-specific dosing with Federated Reinforcement Learning to train models with privacy guarantees. By constructing a well-designed MDP with a reward function constrained by clinical safety, a pharmacokinetic (PK) surrogate model and a federated training protocol with differential privacy, RxAgent consistently delivers state-of-the-art dosing accuracy, therapeutic efficacy and medication compliance, with high safety and privacy assurances. Across all criteria for evaluation, from benchmark comparison and ablation studies, to per-disease breakdowns and pharmacokinetic surrogate accuracy, RxAgent consistently shows the benefits of integrating adaptive RL reasoning with federated privacy protection in an embedded real-time medical device. RxAgent is a step towards the promise of precision medicine at home: not only can devices dispense medications on time, they can reason about patient physiology and learn to improve health outcomes.

References

[1] Osterberg, L., & Blaschke, T. (2005). Adherence to medication. New England Journal of Medicine, 353(5), 487–497.

[2] Cramer, J. A., et al. (1989). Medication compliance and substitution. Epilepsia, 30(4), 523–531.

[3] Checchi, K. D., et al. (2014). Electronic medication packaging devices and medication adherence. JAMA, 312(12), 1237–1247.

[4] Islam, S. R., et al. (2015). The Internet of Things for health care. IEEE Access, 3, 678–708.

[5] Alshammari, H., et al. (2020). IoT-based smart medicine dispenser. Journal of Healthcare Engineering, 2020, 8829171.

[6] Smith, A. J., et al. (2019). Automated pill dispensers for elderly patients. Age and Ageing, 48(3), 329–337.

[7] Turner, M. A., et al. (2018). Paediatric drug development. British Journal of Clinical Pharmacology, 84(5), 963–966.

[8] Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press.

[9] Komorowski, M., et al. (2018). The artificial intelligence clinician learns optimal treatment strategies for sepsis. Nature Medicine, 24(11), 1716–1720.

[10] Prasad, N., et al. (2017). A reinforcement learning approach to weaning of mechanical ventilation. UAI 2017.

[11] Bastani, M. (2014). Model-free intelligent diabetes management using machine learning. Computers in Biology and Medicine, 60, 29–38.

[12] Fox, I., et al. (2020). Deep RL for closed-loop glucose control in T1DM. ML4H Workshop, NeurIPS 2020.

[13] Prasad, N., et al. (2017). RL for weaning mechanical ventilation in ICUs. Proceedings of UAI Conference.

[14] Yauney, G., & Shah, P. (2018). RL with action-derived rewards for chemotherapy dosing. MLHC 2018.

[15] Liu, Y., et al. (2022). Federated RL for clinical decision support. AAAI Workshop on AI for Health.

[16] McMahan, B., et al. (2017). Communication-efficient learning from decentralized data. AISTATS 2017.

[17] Rieke, N., et al. (2020). The future of digital health with federated learning. npj Digital Medicine, 3(1), 119.

[18] Sheller, M. J., et al. (2020). Federated learning in medicine. Scientific Reports, 10, 12598.

[19] Dou, Q., et al. (2021). Federated deep learning for detecting pulmonary nodules. Medical Image Analysis, 73, 102008.

[20] Xu, J., et al. (2021). Federated learning for healthcare informatics. Journal of Healthcare Informatics Research, 5(1), 1–19.

[21] Zhuo, H. H., et al. (2019). Federated reinforcement learning. arXiv:1901.08277.

[22] Nadiger, C., et al. (2019). Application of federated learning to healthcare bandit problems. IEEE Big Data.

[23] Dwork, C., et al. (2006). Calibrating noise to sensitivity in private data analysis. TCC 2006.

[24] Abadi, M., et al. (2016). Deep learning with differential privacy. ACM CCS 2016.

[25] Bonawitz, K., et al. (2017). Practical secure aggregation for privacy-preserving ML. ACM CCS 2017.

[26] Sheiner, L. B., & Ludden, T. M. (1992). Population pharmacokinetics/dynamics. Annual Review of Pharmacology.

[27] Lu, J., et al. (2021). Deep learning for pharmacokinetics. Journal of Pharmaceutical Sciences, 110(1), 154–163.
[28] Norris, L. A., et al. (2022). Transformer-based models for population PK estimation. Bioinformatics, 38(12), 3278–3286.
[29] Qian, Z., et al. (2021). Neural ODEs for personalised cancer dosing. npj Systems Biology and Applications, 7, 38.
[30] Hu, K., et al. (2023). Bayesian deep learning for uncertainty-aware PK dose recommendations. Journal of Biomedical Informatics, 138, 104268.