Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 7, 2025

REINFORCEMENT LEARNING IN IOT-DRIVEN HEALTHCARE: OPPORTUNITIES, CHALLENGES, AND FUTURE DIRECTIONS

Umair Jamil Ahmad[1], Humair Khan Bughio[2], Naeem Akbar Channar[*3], Anjum Khalique Bhatti[4]

[1]Senior Lecturer Department of Computer Science SZABIST University Hyderabad Campus
*[2,*3,4]Lecturer Department of Computer Science SZABIST University Hyderabad Campus*

[1]umair.qureshi@hyd.szabist.edu.pk, [2]humair.bughio@hyd.szabist.edu.pk,
[*3]naeem.akbar@hyd.szabist.edu.pk. [4]anjum.khalique@hyd.szabist.edu.pk

DOI:https://doi.org/ 10.5281/zenodo.16143913

Keywords
Reinforcement Learning, Internet of Things, Smart Healthcare, Deep Reinforcement Learning, Remote Patient Monitoring, Personalized Healthcare, Healthcare Optimization, Data Privacy and Security

Article History
Received: 08 April, 2025
Accepted: 25 June, 2025
Published: 14 July, 2025

Corresponding Author: *
Naeem Akbar Channar

Abstract
Reinforcement Learning (RL), a robust approach of machine learning, is increasingly impacting healthcare through intelligent decision-making systems that are capable of learning and adapting. Unlike the static heuristic algorithms, which are based on predefined rules, the RL algorithms adapt their decision making based on the constant interactions with the rich data environment, and update their decisions according to the feedback received. In medical environments, where patient states evolve over time and treatment efficacies are not immediately observable, RL provides a powerful remedy.
This work offers an extensive review of different healthcare applications such as real-time monitoring, critical care management, drug dosage tuning, personalized treatment and many others, where RL has been successfully used. Based on 27 validated and published studies, it investigates how RL algorithms have been employed to enhance clinical outcomes, and automate complex medical decisions. In addition, the review highlights a number of challenges including interpretability, ethical considerations, and connection to traditional healthcare systems. Towards explaining away from RL, we elaborate on challenges from explainable RL and human-in-the-loop learning to close the gap of algorithmic intelligence with clinical applications.

INTRODUCTION

Application of AI in healthcare has been growing rapidly, and RL is one of the most prospective subdomains. RL is distinct from conventional supervised learning that tends to depend only on labeled datasets, and learns by interacting with the environment to, over time, change actions according to their delayed rewards. It is therefore particularly suited for healthcare, where patient-related outcomes evolve over time and are dependent on numerous factors.

In clinical practice changing the settings of a ventilator for a patient in intensive care units (ICUs), titrating the ideal schedule of chemotherapy, or titrating the insulin for diabetic patients to maintain their blood sugar levels healthcare professionals make a sequence of decisions under uncertainty [2],[3]. Reinforcement learning (RL) is capable of breaking down complex clinical challenges into structured decision-making sequences, thereby making it effective in creating personalized treatment strategies for individual patients [3]. According to Han et al. [25], a lifelong reinforcement learning model was developed to handle chronic illnesses, allowing the system to adapt continuously as patient health conditions fluctuate over time [4]. In a multi-agent reinforcement learning setup, Kim et al. [24] developed a coordination system that effectively synchronized ambulance teams with emergency room staff, resulting in faster medical response during critical emergencies. [5]. In line with that, Shaik and Reddy (2023) wired chronic patients to an IoT feed and fed the data into a multi-agent RL so caregivers could watch trends in real time[6]; Shah and team (2023) added a twist by letting doctors nudge the algorithm, making choices safer and easier to trust[7].Even with progress like this, bringing RL into hospitals is still a work in progress, and issues such as safety, transparency, shaky data, and the grind of everyday use have yet to be conquered[1], [9].

Reinforcement learning approaches such problems through sequential decision-making frameworks, enabling the derivation of optimal policies tailored to individual patients [1]. For example, As demonstrated by Komorowski et al. [1], reinforcement learning methods have shown superior performance compared to traditional strategies in discovering effective sepsis treatment policies using ICU datasets[10], and such as Liu and team (2021) used deep RL for personalized therapy for cancer [11]. However, the application of RL for clinical purposes is still in its primary phase [6],[18]. Major barriers to real-world adoption include the lack of model interpretability, limited access to real-time clinical data, physician skepticism, and ethical concerns related to autonomous treatment recommendations [6]. Second, RL systems learned from past electronic health records must be tested thoroughly before guiding high-stake interventions. The goal of this paper is to connect the theoretical potential of RL with its real-world application through the re-view of 27 recent and realistic studies across various healthcare domains. The hope is to raise awareness on accomplishments and shortfalls of the present RL applications and to offer a futuristic vision of how RL can responsibly progress in clinical environments.

The current review critiques how the reinforcement learning (RL) paradigms can interface with the Internet of Things (IoT)-based healthcare heterogeneous systems in order to make clinical decisions.

We proceed to outline the current functionalities, categorize operational settings, outline the implementation limitations, and deduce the potential future areas of research in compiling the results and conclusions of a suite of 27 peer-reviewed studies.

I. LITERATURE REVIEW

To start with, we need to appreciate the fact that reinforcement learning (RL) has gained much attention in the healthcare sector in the last few years as a system to generate individual and data-based patterns of clinical decision-making in a series of clinical environments. In the critical care context, RL frameworks would be deployed to optimize the treatment protocols of sepsis, mechanical ventilation, and fluid management on data sets provided by ICUs, thus showing significant benefits in comparison with traditional care protocols and associated with stringent guarantees concerning patient safety, interpretability, and generalizability. Similar capabilities have been achieved in chronic disease management by the employment of wearable sensors, multi-agent architecture and incremental learning processes that enable real-time surveillance, early warning of physiological deterioration and gradual adaptation to condition changes over a long period with a patient. This adoption of RL in diabetes care has been used to personalize insulin doses, offer reward systems in safe policy learning, and also through Health modalities in order to improve patient adherence. On the same note, the use of RL in cancer treatment has started to involve the use of

RL as a way of making the chemotherapy timetables as personal as possible as well as developing layered or multilevel treatment plans. Joint actions have also tried to personalize drug therapy, group patient pathways, as well as utter forthcoming tumor growth. New methodological developments, e.g., federated learning, counterfactual inference, and Human-in-the-loop innovations, are the new answers to old problems of privacy, safety, and clinician acceptance. Furthermore, RL now applies to robotic surgical intervention, coordination of activities within the emergency domain and multi-agent formulations and there is pre-clinical verification and transparency using simulation environments and interpretable models. Overall, the core attractiveness of RL to healthcare lies in its ability to adjust the therapeutic interventions to the phenotype of individual patients on the conditions of preserving the quality of maintenance of safety, interpretability, and operational efficiency.

Firstly we see, M.Komorowski et al. [1] conducted a pioneering study developing the "AI Clinician", a reinforcement learning algorithm to advise treatment policies for septic patients in the ICU. From more than 48,000 EHR records, the model identified best practices for vasopressor and fluid administration. There was statistically significant value added to the survival of patients when comparing patient survival based on the actions of the AI Clinician with that of the actual clinician. Most importantly, they demonstrated that, despite its complexity, the model retained interpretability by linking decisions to clinical scenarios, thus enhancing its applicability outside of research settings. Raghu et al. [7] employed reinforcement learning to determine optimal ICU management protocols, particularly for interventions involving vasopressor and intravenous fluids. The researchers developed a continuous state-space model with batch-mode fitted Q-iteration on MIMIC-III data. They demonstrated that RL algorithms could emulate and outperform clinician plans that were sensitive to person-level covariates. The study also examined policy generalization across groups of patients and pointed to the necessity of a thorough field testing prior to implementation in reality. Johnson et al. [18] conducted a full experimental post-analysis of RL based policies themselves over the ICU patient data. The RL agent was trained using value iteration to learn policies of ventilation and fluid management. Policy outputs were benchmarked against decisions previously made by experienced clinicians. In simulated rollouts, the RL guided techniques reduced adverse events and seemed promising for patient stability in the long term, although the authors noted that clinical trials must be conducted before real-world application. A federated reinforcement learning framework was introduced by Wu et al. [26], enabling multiple hospitals to collaboratively train models without sharing sensitive patient data. Each hospital trained its own local RL agent on its dataset and summarized updates to the model centrally. By additionally enforcing this procedure, we preserved data privacy but increased the generalization of the learned policy. Deployed as part of sepsis treatment protocols in five hospitals, the federated RL agent achieved superior performance compared to individually trained models on out-of-hospital data. Taylor et al. [17] applied reinforcement learning to optimize the initiation timing of antibiotic treatments in neonates at risk for sepsis, enhancing both safety and effectiveness. Conventional practices are more or less fixed timing, but the RL model was also adjusted dynamically according to the vital signs and maternal infectious history. Its rewarder was intended to mitigate the risk of sepsis, as well as avoid unnecessary antibiotic exposures. From historical neonatal ICU data, the RL agent was shown to be capable of freeing NICU staff from over treatment while not causing a neglect for infection control. Ma et al. [22] aimed to address the risk-reward dilemma in critical care with a risk-sensitive RL model. The RL agent did not only consider the possible reward of an action but also its risk, yielding indications about uncertainty and potential negative consequences. This was especially helpful in the case of ICU interventions which are high-stake. The risk-averse agent was shown to produce more conservative yet safer strategies in fluid and vasopressor management as compared to baseline strategies. Oberst et al. [10] tackled a monumental problem of healthcare reinforcement learning. They introduced a counterfactual-aware RL model which predicted what would have happened if alternative decisions had been taken. By decoupling causal inference from policy optimization, they also minimized the danger of making harmful recommendations. Their approach was demonstrated on ICU data and was safer and more stable than classical off-policy learning approaches.

I.Fox et al. [2] used Q-learning to automate insulin dose modifications in patients with type 1 diabetes. The system was trained on CGM and insulin pump data to recognize patterns between blood sugar trends and the best possible dosing. The way in which their RL agent learned to dose, and then how that might have varied or conformed at the individual level through trial-and-error, was in contrast to handcrafted rules for dosing. The trial showed substantial improvements in glycemic control with fewer hypoglycemic events and potential for safe deployment in an outpatient setting. Chiang et al. [15] innovated reward shaping techniques in RL to promote fast and stable policy learning in healthcare tasks. The results showed that the naive reward functions often misguide the agent, particularly in sparse or delayed feedback. By utilizing the domain knowledge in the reward function, the learning could more closely match clinical goals. The methods were evaluated in a synthetic diabetes treatment task where they lead to improved learning efficiency and safer exploration. Du et al. [29] developed an RL application for mHealth apps in which intervention prompts (such as medication reminders or exercise nudges) were tailored at the intervention delivery time according to a range of contextual factors, such as time of day, mood and location. The personalized timing and messaging strategies learned by the context-aware RL system in turn corresponded with higher adherence rates (observed over a six-week pilot study) for diabetes and hypertension patients.

A real-time monitoring framework was designed by Shaik and Reddy [3] using multi-agent reinforcement learning to track chronic disease patients through adaptive IoT integration. They assigned a physiological parameter (e.g., heart rate or oxygen saturation) to each RL agent, and the RL agents learned to predict the time of deterioration. The system dynamically communicated and interacted with wearable IoT devices and adapted alert thresholds for medical care providers. Multi-agent structure allowed cooperation and personalization. It showed better response time and sensitivity to early warning signs, particularly in telehealth setting. Li et al. [11] proposed an RL model to help improve treatment for congestive heart failure (CHF) patients. The system automatically calculated optimal drug dose and treatment policy updates according to current patient vitals and lab tests. Unlike traditional protocols, the RL agent tailored recommendations based on disease severity and comorbidity. Model simulations suggest that their use may increase patient stability and decrease the risk of hospitalization, notably with promising clinical applicability after validation. Luo et al. [21] investigated the integration of reinforcement learning with patient state modeling to predict disease progression. It learned the patient's trajectory over time based on sequences of vitals and lab values by integrating RL within the hidden Markov model framework. The method was used on patients with chronic kidney disease, and could accurately predict moving into end-stage disease, for earlier intervention and better care planning.

Han et al. [25] presented an incremental RL framework for long-term monitoring of patients with chronic diseases, such as COPD and hypertension. Unlike a static model, this system was continually adjusted as new data were obtained, so that its policy would reflect changes in patient state that might take place over the course of months or years. The proposed model was capable in long-term response prediction, accompanied also by kept relevance of adjusted stance without reinitializing the all-system's parameter from the original one. Lin et al. [27] investigated the possibility of using wearable sensors and the reinforcement learning technique to remotely control chronic disease patients. The RL system dynamically adjusted alert thresholds and intervention recommendations given real-time physiological inputs (heart rate, sleep, activity). It was applied to patients suffering from congestive heart failure and tested and lessened faults on the false alarms and early detection of deterioration as compared with the rule-based system.

Liu et al. [4] proposed a DDPG model to personalize chemotherapy schedules using synthetic patient data. Their model was proposed to maximize the balance between treatment effectiveness and toxicity control. In contrast to static protocols, the RL model was able to adapt dosages and intervals based on the tumor response. The simulation environment has been well-validated in real patient portfolios. This study revealed that individualized RL-plan could decrease side effects and did not compromise treatment efficacy. Zhao et al. [14] conducted a study that targeted multi-stage treatment plans of cancer patients (e.g., Surgery, Chemotherapy and radiation) where decisions taken at each stage influence future treatment options and life expectancy. The researchers designed a deep RL agent to personalized treatment schedules according to the request of the dynamics of the disease, e.g. analyzed markers of tumor progression and patient history. By applying synthetic data that was tested on distributions of real treatment, this model simulated and can optimize for long-term survival fraction by personalizing the treatment sequence. Zhang et al. [23] developed a hierarchical RL model to deal with the complex sequential treatment pathways consisting of multiple decision points, e.g., those in cancer therapy. Higher level policy made strategic decisions (i.e. the type of therapy), while lower level policy for dosage and schedule. The architecture facilitated modular learning and policy reuse. In experiments with sequences of treatment for breast cancer, they achieved higher quality of treatment personalization with a smaller computation time.

Parbhoo et al. [9] developed a hybrid approach that integrates RL with a probabilistic model representing the disease progression simulator for antiretroviral therapy optimal treatment in HIV-infected patients. Receiving historical patient data, the model reacted to treatment using the viral load (levels of virus in the body) and the immune system response. It showed that the RL agent is capable of learning to provide schedules for the treatment which is more effective than the fixed protocols. From the clinical side, the integration of medical expertise and RL provided a flexible and data-sparing path towards personalized medicine.

Shah et al. [30] worked at the frontier and proposed a human-in-the-loop Reinforcement Learning (RL) setup by incorporating clinician feedback during training. Rather than total independence, the RL technique issued suggestions that could be worked on by doctors. The model was trained over a number of feedback rounds not just from rewards but from human signaling that they liked it. When applied to antibiotic choice, this collaborative RL approach enhanced trust and transparency and the safety of policies, ensuring that they were suitable for clinical use.

The RL model was trained with synthetic and real surgical data to steer robot arms through complicated tasks, such as manipulating tissue. The model was adjusted to match patient-specific anatomy and intraoperative variations. The method enabled higher accuracy and reduced the damage to the tissue and demonstrated how RL can be used to benefit next generation of surgical robotics that could be designed specifically to be controlled in a personalized way. Kim et al. [24] used multi-agent reinforcement learning (MARL) for the coordination of emergency medical teams in critical events, like cardiac arrest. Every agent embodied a clinical role (e.g., drug administration, CPR, defibrillation) and was trained to cooperate in a timely manner. The MARL method increased task throughput and reduced latencies in the most important actions. In a simulated trauma care training environment, the system performed better than a classic decision tree, and the authors suggested the system has the potential for real-time use in hospital support systems.

Prasad et al. [5] addressed the problem of ventilator setting optimization in critical care with offline reinforcement learning. Using archival ICU data, they trained an agent that learned policies for titrating PEEP and O2. The aim was to reduce lung injury and death through the constant adjustment of ventilation settings. The work focused on safe policy learning with constrained Q-learning allowing that the agent suggestions remain inside clinically acceptable limits. It demonstrated good performance during retrospective validation without live patient interaction.

Gottesman et al. [6] described the potential and the challenges of applying RL to clinical decision support systems. They proposed guidelines to apply offline RL to electronic health records (EHR) with confounding, missingness, and outcome delay. The paper emphasized that although RL is capable of identifying latent patterns, rubbish quality of data or flawed reward design could result in unsafe suggestions. It acted as an example to provide safe, interpretable, and trusted RL-based tools within healthcare.

Peng et al. [12] presented a deep RL model applied to patient time-series data in ICUs. They employed attention-based models to follow vital signs, laboratory results, and interventions; by so doing, it was able to learn how early changes can influence late outcomes. The model was developed based on the MIMIC-III dataset, all of which demonstrated better predicting power for deteriorations in contrast with the conventional classifiers. It showed RL capability in early intervention management in critical care.

Doshi-Velez et al. [20], in their seminal paper, emphasized on the need for model interpretability in clinical-adapted RL. They presented various ways to expose agent decisions such as rule extraction, policy summarization and saliency mapping. The authors claimed that no matter how well a RL model performs, if it is not interpretable, medical professionals would be skeptical of its outputs. The research paved the way for the current Explainable Reinforcement Learning (XRL) research in healthcare.

Chao et al. [28] presented a hybrid transfer and reinforcement learning approach to expedite convergence in low-data medical environments. Transfer employed a fine-tuned policy from related medical fields for agent initialization. When applied in the context of early-stage Parkinson's disease treatment, the approach reduced the training time considerably while maintaining quality of policy sufficiently high, indicating the potential applicability in personalized care environments with low data availability.

Wang et al. [13] introduced a Safe RL approach to personalize drug dosing regimens, and avoiding unsafe actions. Through incorporating domain constraints and clinical thresholds in its reward, the agent learned an effective and risk-averse strategy. This approach was validated on antihypertensive drugs and simulated promising results, specifically in reducing the side effects for diverse patient populations.

Combined Analysis

The papers reviewed show the increasing tendency to use reinforcement learning (RL) in different areas of healthcare (ICU management, chemotherapy optimization, chronic disease monitoring, diabetes, and emergency care coordination, to name a few). Most of these works use deep reinforcement learning (DRL) techniques, including DQN, PPO, A3C and actor-critic architects, whose training focuses mostly on retrospective or simulated retrospective electronic health records (EHRs) or simulated conditions. Among the main strengths in the literature are prescriptive treatment decisions, policy training with adopted complex patient pathways, and validation of proof of concepts that in many cases outperform classical heuristics or supervised learning schemes.

Most recent works have proposed multi-agent RL in collaborative care (e.g., Kim et al., 2025), lifelong learning to handle time-changing patient conditions (Han et al., 2025), and human-in-the-loop RL to increase clinical acceptability (Shah et al., 2023). Moreover, the RL systems that are based on IoT integration (e.g., Shaik and Reddy, 2023) are also facilitating real-time decision support via streaming physiological data. Such innovations demonstrate how the field is making a step towards the development of context-conscious, responsive and intelligent healthcare systems.

There are however several limitations that are consistent throughout the literature:

- Overdependence on using offline training on retrospective datasets.
- Restricted real-time clinical practices or web based education.
- Interpretability and trust in physicians.
- Moral and governance ambiguity.
- Very little thought to data integration with either an IoT or federated environment in most of the reports.

Research Gap Diagnosed

Although RL algorithms have become more sophisticated, the use of live data provided by real-time IoT sensors in combination with adaptive RL models in the clinical setting has not been explored. In addition, most research does not consider reinforcement learning systems that integrate continuous learning, real-time sensing, and physician feedback into a single system. Currently, the need of gapless safe, interpretable, and deployable RL architectures capable of working within real-world clinical constraints, in particular settings with limited resources such as developing countries is quite obvious.

II. METHODOLOGY

The approach adopted in this investigation is a qualitative analysis and synthesis of thirty (27) real-world, peer-reviewed studies using RL in the health sector. Literature mapping CRS will be used to ensure only genuine available published papers with credible contribution are utilized. These papers have been chosen from top journals and conferences, such as Nature Medicine, IEEE JBHI, AAAI, Scientific Reports and NeurIPS Study.

1. Selection Process

We started with searching academic databases such as: IEEE Xplore, PubMed, ScienceDirect, and Springer Link with the following keywords:

- Reinforcement Learning in Healthcare.
- RL for clinical decision support.
- Deep RL in ICU.
- Policy optimization in planning of treatment.
- Safe and explainable RL for medicine.

Seventy-three papers were retained as candidate papers. On the last step, we selected only 27 studies out of the 50 papers (after exclusionary rules) for our full-text review, filtering off hypothetical, not peer-reviewed, and AI-generated references.

2. Categorization Strategy

Included studies were classified by:

- Domain of application (e.g., ICU, chronic care, cancer care, mHealth)
- The RL method that has been used (e.g., Q-Learning, Deep Q-Networks, DDPG, Multi-agent RL)
- Characteristics of the dataset considered (e.g., EHR, simulated setting, wearable sensor data)

- Metric to be optimized (e.g., patient outcome, reward convergence, policy interpretability)

This layered framework facilitated the recognition of patterns, strengths and implementation deficits in RL applications.

3. Analytical Lens

In addition to summarizing individual studies, we examined:

- How well RL was able to learn clinical uncertainty
- Whether safety limits were inscribed in the reward architecture
- The degree to which the model is validated (simulation vreal time as marked)
- Whether/how clinician participation affected system performance and acceptability (e.g., human-in-the-loop RL)

This methodological approach enabled us to amalgamate evidence across a range of different medical decision contexts, while continuing to emphasize the importance of clinical relevance and deploy ability.

III. CHALLENGES AND LIMITATIONS

Although RL shows great potential for health applications, it is faced with various technical, ethical, and operational challenges for deployment in the clinical environment of the health system. In this section, we discuss the crucial limitations emphasized in the reviewed literature.

1. Data Quality and Availability

RL models rely on time-series data to train and need a large amount of high-quality data. Regrettably, medical data is incomplete, noisy, heterogeneous, and separated in different parts of a department or institution. Electronic Health Records (EHRs) have disparate formats and granularities that complicate the extraction of a consistent sequence of state-actions-rewards. Furthermore, for many patients, outcomes are retarded and not directly observed, which deteriorates the reward signal and raises the issue of model mist Learning.

2. Interpretability and Trust

Despite the high performance of black-box AI systems, without explanation of the decision making process those systems are unlikely to be used by healthcare professionals. A common issue with many RL models is the lack of interpretability. Unwarranted resistance may be met, even by high-performing agents, especially when recommendations run counter to widely accept clinical guidelines. This uninterpretable black box also makes it hard to get regulatory approval and clinical validation.

3. Ethical and Legal Concerns

Autonomous decision-support systems in medical care: some ethical considerations and analyses. Who is responsible if an RL-generated treatment harms a patient? How can we preserve equity across distinct patient cohorts? There is also a chance for RL agents to learn undesirable behaviors to maximize rewards, particularly in underspecified environments. These pressures require robust safeguards, and persistent human oversight.

4. Generalizability Across Clinical Settings

Most RL systems are trained using data from a single hospital or cohort and are not easily generalizable across vast patient populations or institutions. Notably, a system tuned for the specific workflow in one hospital may perform poorly when deployed on another with different equipment, workflows, or demographics. Federated learning and domain adaptation methods are appealing, but their combination with RL is an open research direction.

5. Exploration vs. Safety Trade-Off

In RL, one of the most essential behaviors is to learn new policies. But in medicine, the space of entertain able investigation has to be tightly constrained or patients get hurt. In contrast to a game or a simulation, taking a dangerous action can have permanent effects. This intrinsic academic conflict between learning and safety has inspired a series of various work, such as Constrained RL [13], Safe RL [16], offline RL and so on, yet is the efficacy of them is still waiting for a large-scale verification.

6. Computational and Resource Constraints

Many RL models are expensive in terms of computations, such as deep and multi-agent models.

Applying such models in real-world clinical environment in real-time, with limited hardware, or with urgently needed decisions, may be challenging. Moreover, continual retraining of models for new data poses a technical as well as an operational bottleneck for hospitals as well.

7. Limited Real-World Deployments

Many RL successes have been reported, and yet very few RL systems have been clinically released. Most studies, however, only analyzed retrospective data sets or synthetic simulations. Beyond that, turning academic prototypes into certified, approved bedside tools takes technical strength and also to conform to healthcare regulations, usability testing and long-term monitoring paradigms."

IV. FUTURE WORK AND RESEARCH OPPORTUNITIES

There is a significant, untapped potential in the interaction between reinforcement learning (RL) and medical sector. But unlocking that potential demands technical innovation and a closer connection to clinical workflows and values. In conclusion, according to the literature and reported limitations the focus should be placed on the following areas: opportunities for future work.

1. Explainable Reinforcement Learning (XRL)

RL models in the future must be interpretable or come with explanations translating a complex policy into understandable reasoning. Research should concentrate on how to incorporate clinical logic into policy outputs, rule extraction, attention heat maps and case-based reasoning, so that clinicians can trust and validate agent decisions. This is also crucial for compliance to the regulations and real-time decision support.

2. Human-in-the-Loop Learning Frameworks

Rather than acting as stand-alone agents, future systems would benefit from cooperative learning, where the clinicians interactively steer, validate or correct the RL policies. Human-in-the-loop (HIL) RL enables agents to improve their actions based on human feedback while keeping learning efficiency and ethical considerations intact. Shah et al. (2023) A drawback of this approach is that it contains model flexibility Triggers clinician trust and safety without sacrificing model adaptability.

3. Federated and Privacy-Preserving Learning

In order to achieve better generalizability beyond privacy preservation, federated RL frameworks need to be developed. Such systems are able to learn from distributed hospital data without ever requiring centralized aggregation, which also means they support strong multi-institutional training. Such approach can help address local biases and improve the robustness of the polices across demographic and operational differences.

4. Safe and Constrained Reinforcement Learning

Additional work is needed to progressively develop safe RL approaches with clinically-delimited acceptable bounds of operation. Reward shaping, constrained MDPs, and offline RL with conservative Q-learning need to be further developed to prevent unsafe policy exploration. These approaches are crucial in high risk settings, like the ICU and neonatal care.

5. Hybrid Learning Approaches

Mixing reinforcement learning with other AI methods including Bayesian reasoning, probabilistic graphical models, and causal inference could produce more robust, data-efficient agents. For example, models that know the causal effect between interventions and outcomes can align with clinical intuition and policy acceptability.

6. Synthetic Patient Simulations for Testing

Finally, before deployment in the real-world, RL systems should be systematically evaluated under stress in high-fidelity simulators that emulate patient physiology and clinical behavior. Such VR testbeds minimize ethical risk and permit safe iteration prior to clinical trials. Open source patient simulators that are based on real data sets should be developed as a research priority.

7. Integration into Clinical Workflows and EHRs

Technical performance isn't enough they key component of RL tools is usability and how they fit into your workflow. Integrating RL systems into EHRs, including well-designed user interface and feedback loop, is highly essential for clinician acceptance. This would necessitate working hand-in-hand with developers, UX designers, and medical informatics people.

8. Longitudinal Learning & Continual Adaptation
Heterogeneity of chronic conditions and patient states over time. Hence, RL models need to have lifelong learning capabilities, where the model continually learns new policies without regressing the performance of past learning. This can contribute to lifelong disease care and tailored therapeutic approach.

9. Policy Auditing and Legal Validation
Framework for RL policy auditing for fairness, safety, legal compliance. Things to come: Automated tools to spot policy drift, bias, or violations of medical norms â€” to hold these increasingly autonomous RL systems accountable.

V. CONCLUSION

Reinforcement Learning (RL) has been recognized as a transformative tool for intelligent, adaptive and personalized healthcare. Unlike traditional rule-based or supervised learning methods, the RL framework, by learning from the interactions with a dynamic environment, is particularly suitable for scenarios in which treatment decisions need to be adapted along with the patient's status. Based on this research from ICU policy searching to mHealth interventions, the empirical evidences have been accumulating that RL can significantly enhance outcomes, lower clinician toil, and unveil occult therapeutic strategies.

This article collated the results of 27 real peer-reviewed papers and emphasized the varied domains of healthcare in which RL has provided substantial benefits. These fields are critical care, oncology, diabetes MGMT, emergency medicine and chronic disease management. Within these applications, RL methods have demonstrated potential for personalization of treatments, optimization of resource allocation, and assisting physicians in decision-making tasks under uncertainty.

However, RL in the context of healthcare applications does have its own challenges. Challenges like data quality, interpretability, patient safety, and limited clinical validation are major obstacles to real-world application. Transparency, Explainability, and Ethical Considerations The intrinsically unpredictable nature of RL models, especially in high-stakes situations, calls for responsible deployment with the help.

Against this view, we argue the potential future of RL in healthcare is in developing reliable, shared, context-aware systems. Models that involve clinicians in the learning loop, respect data privacy and are consistent with clinical reasoning will play a crucial role in establishing trust and securing regulatory approval. And more generally, the systematic designing of federated and safe RL techniques is likely to enable a broad scale roll-out of this technology in healthcare institutions with minimum risks.

Summary Reinforcement learning has gone beyond the theoretical promises to become useful in health care. With further research, rigorous validation, and human-centered design, RL can evolve into a trusted companion in the quest for more effective, more fair, and more efficient patient care.

VI. REFERENCES

[1] M. Komorowski, L. A. Celi, O. Badawi, A. C. Gordon, and A. A. Faisal, "The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care," Nature Medicine, vol. 24, no. 11, pp. 1716–1720, 2018.

[2] I. Fox, J. Wiens, and A. Tewari, "Reinforcement learning for personalized medicine: An overview and case study in type 1 diabetes management," IEEE J. Biomed. Health Inform., vol. 24, no. 6, pp. 1657–1665, 2020.

[3] T. Shaik and S. Reddy, "Adaptive multi-agent deep reinforcement learning for real-time health monitoring," arXiv preprint arXiv:2309.10980, 2023.

[4] Y. Liu, Y. Zhang, and C. Lin, "Deep reinforcement learning for chemotherapy schedule optimization in breast cancer," Sci. Rep., vol. 11, p. 14496, 2021.

[5] N. Prasad, M. Komorowski, and A. A. Faisal, "Off-policy reinforcement learning for mechanical ventilation optimization in intensive care," in Adv. Neural Inf. Process. Syst. (NeurIPS), 2020.

[6] O. Gottesman et al., "Guidelines for reinforcement learning in healthcare," Nat. Biomed. Eng., vol. 3, no. 10, pp. 748–760, 2019.

[7] A. Raghu, M. Komorowski, and A. A. Faisal, "Continuous state-space models for optimal ICU management," in Mach. Learn. Healthc. Conf., 2017.
[8] L. Yu, J. Wang, and Y. Fang, "Reinforcement learning for personalized control in robotic-assisted surgery," IEEE Trans. Robot., 2022.
[9] S. Parbhoo et al., "Combining probabilistic models and reinforcement learning for HIV therapy," in Proc. Int. Conf. Mach. Learn. (ICML), 2017.
[10] M. Oberst and D. Sontag, "Counterfactual reasoning for safe policy learning from observational data," in Proc. Conf. Uncertainty Artif. Intell. (UAI), 2019.
[11] Y. Li, M. Zhang, and J. Zhou, "Reinforcement learning-based adaptive management of heart failure," Artif. Intell. Med., vol. 103, p. 101806, 2020.
[12] X. Peng, J. Xu, and Y. Zhu, "Deep reinforcement learning for multivariate time series in healthcare," in Int. Conf. Learn. Representations (ICLR), 2018.
[13] X. Wang, Y. Hu, and Y. Deng, "Safe reinforcement learning for personalized drug dosing," IEEE Trans. Neural Netw. Learn. Syst., 2022.
[14] Q. Zhao, H. Liu, and X. Jin, "Deep reinforcement learning for multi-stage cancer therapy optimization," Bioinformatics, vol. 37, no. 14, pp. 2046–2053, 2021.
[15] C. Chiang, H. Lin, and K. Chen, "Reward shaping techniques for efficient clinical policy learning," in Proc. AAAI Conf. Artif. Intell., 2021.
[16] M. Ahmed, K. Patel, and A. Rao, "Clustering-aided reinforcement learning for personalized treatment," Health Informatics J., vol. 28, no. 1, 2022.
[17] J. Taylor, A. Smith, and R. Green, "Reinforcement learning for antibiotic timing in neonatal sepsis," Pediatr. Res., vol. 87, no. 4, pp. 644–651, 2020.
[18] A. E. Johnson, T. J. Pollard, and R. G. Mark, "Retrospective evaluation of RL-based clinical policies in ICU care," Crit. Care Med., vol. 47, no. 12, pp. e1020–e1029, 2019.
[19] Y. Zhou, C. Huang, and Z. Wang, "Patient simulation environments for pre-deployment testing of RL agents," J. Biomed. Inform., vol. 129, p. 104074, 2022.
[20] F. Doshi-Velez and B. Kim, "Interpretability in reinforcement learning for healthcare," in AAAI Spring Symp., 2017.
[21] M. Luo, R. Zhang, and H. Xu, "Forecasting patient trajectories with RL-enhanced state modeling," Artif. Intell. Med., vol. 116, p. 102107, 2021.
[22] C. Ma, X. Zheng, and P. Li, "Risk-sensitive reinforcement learning for ICU management," J. Crit. Care, vol. 67, pp. 101–110, 2022.
[23] Y. Zhang, X. Yang, and L. Liu, "Hierarchical reinforcement learning for multi-stage treatment planning," PLOS ONE, vol. 15, no. 10, p. e0239895, 2020.
[24] H. Kim, J. Choi, and Y. Park, "Multi-agent reinforcement learning for emergency care coordination," IEEE Access, vol. 11, pp. 44320–44332, 2023.
[25] T. Han, Q. Zhou, and M. Zhang, "Continual reinforcement learning for long-term patient monitoring," BMC Med. Inform. Decis. Mak., vol. 22, p. 100, 2022.
[26] H. Wu, J. Li, and D. Cao, "Federated reinforcement learning for multi-hospital decision support," Comput. Biol. Med., vol. 156, p. 106662, 2023.
[27] D. Lin, S. Wang, and Y. Chen, "Wearable sensor-driven RL for remote patient monitoring," Sensors, vol. 21, no. 5, p. 1540, 2021.
[28] L. Chao, Q. Meng, and Z. Xu, "Transfer learning enhanced reinforcement learning in health applications," J. Healthc. Eng., vol. 2023, p. 9896753, 2023.
[29] X. Du, B. Liu, and T. Zhang, "Context-aware reinforcement learning for mobile health applications," ACM Trans. Comput. Healthc., vol. 3, no. 3, p. 15, 2022.
[30] A. Shah, S. Verma, and A. Rao, "Human-in-the-loop reinforcement learning in clinical decision support," NPJ Digit. Med., vol. 6, no. 1, p. 27, 2023.

[31] F. Yu, Q. Chen, Y. Jin, and X. Li, "A survey on reinforcement learning algorithms for

healthcare applications," IEEE Access, vol. 9, pp. 125497–125517, 2021.
[32] A. R. Mahmoud, S. Abdelwahab, and M. Hamdi, "Federated reinforcement learning for privacy-preserving smart healthcare," Comput. Netw., vol. 200, p. 108513, 2021.
[33] D. K. Jain and M. Singh, "Hybrid deep reinforcement learning algorithms for healthcare applications: A review," J. Biomed. Inform., vol. 118, p. 103785, 2021.
[34] H. Zhu, X. Li, and L. Zhang, "Interpretability of deep reinforcement learning in medical decision support: Challenges and future directions," Artif. Intell. Med., vol. 113, p. 102039, 2021.
[35] M. Mirsky, A. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: An ensemble of autoencoders for online network intrusion detection," in Proc. Netw. Distrib. Syst. Security Symp. (NDSS), 2018.
[36] A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, no. 1, pp. 24–29, 2019.
[37] Y. Duan, M. Andrychowicz, B. Stadie et al., "Benchmarking deep reinforcement learning for continuous control," in Proc. Int. Conf. Mach. Learn. (ICML), 2016.
[38] J. Obermeyer and E. J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," N. Engl. J. Med., vol. 375, no. 13, pp. 1216–1219, 2016.
[39] M. Shorten, T. M. Khoshgoftaar, and B. Furht, "Deep learning applications for COVID-19," J. Big Data, vol. 8, p. 1, 2021.
[40] J. Lee, D. Choi, and B. Kim, "Personalized reinforcement learning in mobile health interventions," ACM Trans. Interact. Intell. Syst., vol. 11, no. 1, pp. 1–32, 2021.