SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 5, (2026), pp. 597-606
Open Access. Online & Print

DOI: 10.5281/zenodo.20767337

DEEP REINFORCEMENT LEARNING IN ONCOLOGY: ADVANCES IN CANCER IMAGING, RADIOTHERAPY, AND PERSONALIZED TREATMENT

Dr. Ohmini Krishnamurthy Rajendran*

Consultant, MBBS, MD Radiodiagnosis, KIMS Hospital and Research Centre
Krishna Rajendra Road, Parvathipuram, Vishweshwarapura, Basavanagudi, Bengaluru, Karnataka 560004
Email: jsmnk4@gmail.com

Received: 13/12/2025
Accepted: 28/02/2026

Corresponding Author: Dr. Ohmini Krishnamurthy Rajendran (jsmnk4@gmail.com)

ABSTRACT

Reinforcement Learning (RL) enables computers to learn dynamic decision making methods via trial and error, to maximize numerical reward signals. Its roots lie in early work across many fields, but recent advances have improved its capabilities greatly. Cancer is complex and dynamic and hence RL has a lot of potential in the field of oncology especially in imaging where it can improve accuracy and efficiency . In this study, we review RL in cancer and briefly introduce RL algorithms and categories. It then discusses applications of RL in cancer including in imaging and radiation. The report concludes with current challenges and future perspectives, emphasizing RL's promise to personalize cancer diagnosis and therapy.

KEYWORDS: Reinforcement Learning, Oncology, Cancer, Review.

1. INTRODUCTION

Reinforcement Learning (RL) has transformed artificial intelligence by enabling computers to learn dynamic decision making techniques through trial and error. Machines can learn from environment and experience (similar to biological principles) and improve the efficiency of computer science and engineering. In the past few years, deep reinforcement learning has been successful by incorporating deep learning models into reinforcement learning algorithms. OpenAI has popularized RL by solving its main problems, finding new solutions, and providing learning resources [1].

RL is a result of cybernetics, statistics, psychology, neurology, and computer science. Recent advances in theory and technology for RL have made it more useful in natural language processing and computer vision . Adoption has lagged in real-world challenges. The casting of experimental RL setups is different from the ill-defined reality of real-world systems. Few attempts have been made to use these algorithms in difficult fields such as medical imaging and they are still unexplored [2].

Cancer is an important but relatively unexplored RL option in these complex areas. Cancer is one of the leading causes of death globally. Abnormal cell division and spreading which invade others can occur in almost every organ or tissue in the body. Globally, cancer incidence and death are on the rise. The most common cancers are breast, lung, colorectal, and prostate. The treatment of cancer is still one of the greatest challenges to oncology. Depending on the malignancy and the patient, surgery, radiation, immunotherapy, chemotherapy, stem cell transplantation, etc. are used [3].

Medical imaging has a key role in the cancer care continuum from diagnosis and characterisation to real-time monitoring. Its ability to offer low to no invasive tissue accessibility and function over time and size ranges is critical to biological and pathological processes. RL has the potential to improve accuracy and efficiency in cancer imaging. RL algorithms are very good at learning good cancer decision-making techniques from complex data. RL can process large amounts of imaging data and patient specific information, and has the potential to improve treatment planning and real-time monitoring. The changing nature of cancer demands that therapy is flexible and individualized, a skill useful in oncology. Another benefit of RL is that it can deal with delayed effects even if the relationship between activities, such as treatment decisions, and patient outcomes is unknown. This review summarizes the literature on RL in cancer imaging with focus on recent advances, problems and future prospects [4].

2. REINFORCEMENT LEARNING

A. Basics

An RL consists of agents and environments. The agent learns by interacting with the environment and dynamically maps situations to actions to optimize reward signals. Rather than being told what to do, the agent has to try a number of actions to find the most rewarding one. Actions affect future inputs to the system. These are the most typical aspects of RL problems. The closed loop system is shown in Fig. 1 [5].

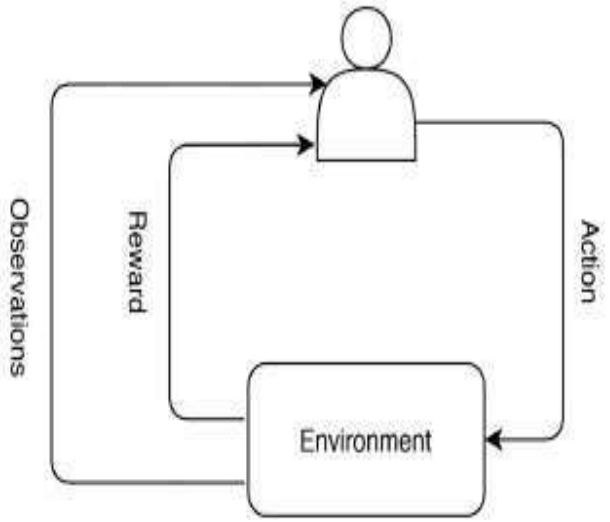

Fig. 1. Agents and environments interaction in RL.

RL is neither supervised nor unsupervised as it maximizes a reward signal from its own experience rather than finding a hidden structure . Unsupervised learning is about finding hidden patterns or structures in the data. Supervised learning is about learning from labeled data. RL agents, on the other hand, learn by interacting with their environment and receiving rewards or penalties [6]. Self-supervised learning tasks do not use this feedback, but rather implicit data supervision. A separate RL issue is that of exploration versus exploitation. To maximize reward, an agent must use past successes and explore new possibilities. This balance is critical and has been extensively studied, as focusing on one results in suboptimal outcomes. RL is different from supervised and unsupervised learning since this problem does not exist. It is seen as a third machine learning paradigm (see Fig. 2), besides supervised, unsupervised and maybe more paradigms [7].

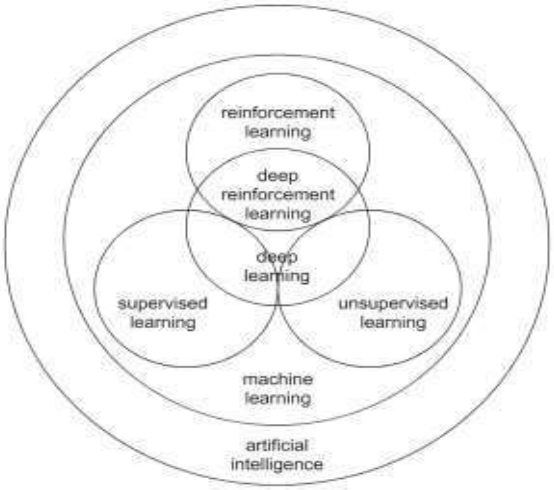

Fig. 2. Machine learning, deep learning, unsupervised, supervised, RL connections.

The classical theory of RL trains an agent to operate in an MDP environment where the rewards depend only on the previous state and action. An MDP is specified by the 5-tuple {S,A,T,R,γ}. The agent is given information about the current state st ∈ S. At each time step t the agent interacts with the environment by taking an action at ∈ A. Agents can select S and A states or actions. After taking an action at ∈ A, the agent will move to a state st+1 ∈ S with a probability given by a transition probability function T(st+1|st, at). Thus, the reward function, R, is the reward rt+1 = R(st, st+1) of taking the action at in state st and transitioning to state st+1. A trajectory τ is a complete sequence of states and actions, e.g. $(s_0, a_0, r_1, s_1, a_1, r_2, \dots)$. An episode occurs when a trajectory goes to infinity and meets a termination condition or limited length. Also, the future reward care of an agent is determined by a discount factor γ ∈ [0, 1][8].

An effective policy π directs the agent to choose and react adaptively according to its current condition. We evaluate this approach using the state value function V \pi (s). It is the expected return from the current state s under policy pi . It is recursively defined by Bellman's expectation:

$$V^{\pi}(s) = \mathbb{E}[r_{t+1} + \gamma V^{\pi}(s_{t+1})] \quad (1)$$

rt+1 = immediate reward from statest action Equation 1 is recursive because the actor wants to maximize the cumulative reward[9].

Expectation term can be calculated by using policy, transition probability and prior return as follows:

$$V^{\pi}(s) = \sum_{a_t \in A} \pi(a_t|s_t) \sum_{s_{t+1} \in S} T(s_{t+1}|s_t, a_t) \cdot [R(s_t, s_{t+1}) + \gamma V^{\pi}(s_{t+1})] \quad (2)$$

T(s_{t+1}|s_t, a_t) is the probability of transitioning from state s_t to s_{t+1} after action a_t, and R(s_t, s_{t+1}) is the immediate reward. In Equation 2[10], V π (s) is calculated recursively by summing all potential actions at and future states st+1 .

The action-value function Q\pi^*(s_t, a_t) is an extension of the state-value function V \pi(s). It is the expected return starting from state s_t and taking action a_t. In RL, Q\^{}pi^{}*\{st, at\} is the expected cumulative reward for starting in state st, taking action at and following optimum policy pi^{}* [11]. The fundamental objective in MDP solution is the optimal policy $\pi ^*$ such that $V_\pi(s)$ is maximized for every policy π and state $s\in S$. Define the optimal state-action value function

$$Q^{\pi^*}(s_t, a_t) = \sum_{s_{t+1} \in S} T(s_{t+1}|s_t, a_t) \cdot [R(s_t, s_{t+1}) + \gamma V^{\pi^*}(s_{t+1})] \quad (3)$$

RL uses the Q-function to learn and update rules so as to maximize cumulative rewards. In general, it measures how good an agent can be in selecting action in state st[12].

Once RL has been introduced, it is possible to compute an optimal policy for an MDP using model-based and model-free RL. The model is the environment. The primary difference is whether the model is built to actively mimic and describe the world or to physically interact with it and track rewards[13].

B. Model-based RL

RL model-based methods rely on world dynamics. They calculate suitable MDP policies using learned prior knowledge of world dynamics without interacting with the environment. Such approaches use transition models to replicate the manner in which activities from one state result in another. Value iteration is used to maximize state value functions to find the optimal policy. Transition models specify the mapping of actions from state s to the next state. Policy search methods directly optimize policy parameters. Value functions are used to select actions by maximizing the value of each state. The return functions accumulate rewards or penalties over episodes, which is important for the convergence and feasibility of the model but hard to describe and model [14].

C. Model-free RL

Model-free RL agents learn by trial-and-error in their interaction with the environment. There are

value-based, policy gradient, and actor-critic model-free RLs. The methods will be described below.

These RL algorithms can also be used with deep learning to build deep reinforcement learning. DRL approximates value function and policy using neural networks. DRL can learn from delayed feedback and take sequential decisions. It is hard to backpropagate the gradient of the reward to prior actions without modeling the joint distribution over choices and states, since agent decisions affect the environment. So DRL can handle non-differentiable metrics well. The next sections discuss DRL algorithms and their applications [15].

1. **Value-based RL:** Value-based RL learns values for states or functions to find the best action policy. Q-learning is usually value-based. The Q-learning rule updates Q-values based on seen and predicted future rewards as follows with a learning rate λ:

$$Q(s_t, a_t) = Q(s_t, a_t) + \lambda\delta_t \qquad (4)$$

Temporal difference error (δt) is the difference between the predicted Q-value and the reward plus the estimated best future value. This iterative process enables the agent to converge to the optimal Q-values, maximize the expected cumulative reward, and improve the decision-making [16].

Then Deep Q-Network (DQN) was proposed. It learns rules from high dimensional inputs using Q learning and deep neural network. Experience replay and a fixed target network stabilize learning of the Q-function in DQN [9]. The experiences of the agent at each iteration are stored in the replay memory buffer and mini-batches of experiences are used for updating the network weights based on the Bellman equation. This greatly reduced sample correlations, thus increasing robustness [17]. However, important transitions are ignored by the replay buffer. To address this limitation and reduce the correlation between current and target Q values, a target Q value network was implemented. Double DQN employed two networks for selection and evaluation, minimizing overestimations, improving performance. Fig. shows the general architecture. 3.

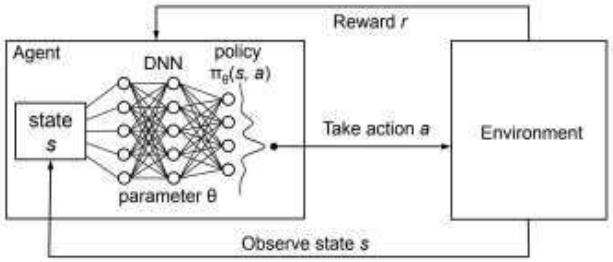

Fig. 3. Deep Q-Network agent-environment interaction.

2. **Policy-based RL:** Policy gradient methods find the optimal policy G(π) → max_π, by optimizing the discounted expected reward, without using MDP information. For large or infinite state-action spaces it is useful to work directly with the objective function. These methods parameterize the policy using approximate estimates of the gradient:

$$\mathcal{G}^{\theta}(\pi) = \mathbb{E}_{T_\tau} \sum_{t=0} log(\pi_\theta(a_t|s_t))\gamma^t \mathcal{R} \qquad (5)$$

R is total cumulative return . θ is an updated parameter .

3. **Actor-critic RL:** Actor-critic methods combine policy gradient and value-based techniques to overcome their weaknesses. However policy gradient methods generally converge to local maxima and have high variance. Value-based methods may not converge. Actor-critic algorithms employ policy gradients to determine actions and temporal difference learning to evaluate the quality of actions, making it possible to make efficient updates for continuous action spaces [18]. The policy gradient is improved using a value-based critic $\Gamma=Q_\pi(s_t, a_t)$ as follows:

$$\mathcal{G}^{\theta}(\pi) = \mathbb{E}_{T_\tau} \sum_{t=0} log(\pi_\theta(a_t|s_t))\Gamma_t \qquad (6)$$

The advanced actor-critic algorithms A2C and A3C share a parallel architecture. A2C updates a shared policy across workers synchronously, which makes a faster convergence. In A3C, workers can update independently, which may speed up A2C convergence by having different update times.

Figure 4 summarizes model-free and model-based reinforcement learning methods. Such algorithms are the basis of several applications in cancer, as described in the next section [19].

Fig. 4. A partial taxonomy of RL algorithms. DQN is Deep Q-Network, SARSA is state-action-reward-state-action algorithm, and AC is actor-critic.

3. APPLICATIONS OF RL IN ONCOLOGY

Several RL approaches used in oncology will be discussed below. The approaches are further classified into radiology, radiotherapy and other uses. Table I shows the applicability of the papers. It categorizes RL algorithms as discussed in the previous section [20].

A. Radiology

Screening for cancer looks for early warning signs when there are no symptoms. Finding cancer early is important to getting the best treatment. The focus of many studies has been on breast cancer screening because it is a common type of cancer. Wang et al. proposed a multimodal framework for the classification of breast nodules as benign or cancer using ultrasound modalities. It uses a multitask learning framework with weight sharing and RL to learn correct modality weighting automatically. The model-free REINFORCE algorithm was used to estimate the gradient of the predicted cumulative reward with respect to policy network parameters using Monte Carlo estimation. The proposed model obtained better results with different ultrasound modalities. This showed that the model is able to learn automatically the best weighting of each modality. This showed that all modalities are necessary for the correct diagnosis.

Yala et al. proposed an RL framework for risk based breast cancer screening policy. Screening was intended as a sequential decision-making process to enhance early detection and lower screening costs. The technological implementation was a multi-layer perceptron with state as risk assessment, actions as screening follow-up suggestions and a reward based on early detection gains and screening costs. We trained the network using an RL technique called Envelope Q-learning which balances several goals (see Section II-C1, Equation 4). They found that the model outperformed the standards of screening in early detection per screening cost vs yearly screening.

TABLE I: A Summary of Applications of Rl in Oncology.

Application Area	Specific Application	RL Category	Algorithm Used
Radiology	Classification of breast cancer	On-policy, Policy Gradient	REINFORCE
Radiology	Screening for breast cancer	Off-policy, Value-based	Envelope Q-learning
Radiology	Diagnosis of breast cancer	Value-based, On-policy	SARSA
Radiology	Skin cancer diagnosis	Value-based, Off-policy	DQN
Radiology	Liver cancer detection	Actor-Critic, On-policy	Soft Actor-Critic
Radiotherapy	Prostate cancer treatment planning	Value-based, Off-policy	Q-learning + VTPN
Radiotherapy	Treatment planning for prostate cancer	Value-based, Off-policy	Q-learning + HieVTPN
Radiotherapy	Clinical decision support	Off-policy, Value-based	Quantum DQN
Radiotherapy	Beam orientation for biopsy planning	Value-based, Off-policy	GCN + DQN
Radiotherapy	Needle placement for biopsy	Off-policy, Value-based	DQN
Chemotherapy	Tailored chemotherapy therapy	Value-based, On-policy	SARSA

Kolla et al. also proposed a breast cancer diagnostic method based on a deep neural network. They constructed a teacher-student model trained with the deep neural network architecture and distillation of knowledge learned from the instructor. A veteran instructor is able to steer a student through difficult problems with ease, just as a mentor might teach a student detective how to find hidden evidence. They used the SARSA algorithm (state-action-reward-state-action), an RL approach, to maximize model student learning. Student model was computationally more efficient than instructor model and had comparable accuracy.

Beyond breast cancer, Barata et al. applied RL to incorporate human preferences into AI-based skin cancer decision support to improve AI performance. The DQN model used a one dimensional state vector with multiclass probabilities and features of the supervised learning model. The action space consisted of diagnostic or treatment decisions and a reward table constructed by clinical experts. It was designed to clarify incentives by providing clinician choices. They demonstrate the importance of human knowledge in decision making. This approach increased the model's sensitivity and reduced the overconfidence in wrong predictions.

Xu et al. proposed a teacher-student RL framework for liver tumor identification from non-enhanced liver tumor magnetic resonance images. The mechanism involves understanding the teacher network's decision making, transferring pixel level information to the student network, and evaluating the student network's actions based on previous performance. This framework outperformed the existing approaches. This suggests that the teacher-student RL architecture can accurately diagnose liver tumors without the need for contrast chemicals.

B. Radiotherapy

Radiation uses high doses of X-rays or other energy to kill cancer cells. All tumors and metastases are skin-beamed. Several publications dealt with

radiation dosage and beam direction therapy planning.

Shen et al. applied Q-learning (see Section II-C1) to optimize prostate cancer intensity-modulated radiation therapy treatment plans with current dose-volume histogram weights. The virtual treatment planner network, a deep neural network trained with DRL, could autonomously change the treatment plan parameters and improve the quality of the plan. This approach resulted in a significant improvement in plan and dose-volume histogram metrics. This proved that advanced DRL methods can improve the radiation planning of prostate cancer. Shen et al. developed HieVTPN, a hierarchical DRL framework to scale and improve automated radiation therapy treatment planning [21].

Another DRL method was developed by Niraula et al. In particular, they have validated a DRL framework for clinical decision support in adaptive radiotherapy for patients with non-small cell lung cancer. This novel approach to clinical decision making is improved by quantum inspired techniques. They used the DQN model to modify treatment strategies according to patient reactions and changing situations. The radiation dosage was the decision-making policy and the transition functions of the neural networks were for different patient characteristics and outcome estimators. They demonstrated a 10% improvement in the efficacy of clinical radiotherapy decision making compared to current techniques.

Kafaei et al. proposed a new technique, a graph neural network-based DQN method, for Cyberknife treatment planning beam direction. They employed graph neural networks to identify the optimal beam subset and sequence for the beam orientation optimization. To reduce the distance between robotic arms, avoid picking closely spaced beams and to introduce beam ratings based on beam intensity, a complex reward mechanism was implemented. We trained the network on the RL learning environment, an MDP with states as graphs with node properties, actions, transitions and rewards, and used prioritized experience replay to stabilize training. The proposed method reduced the treatment time by 35% with respect to the classic clinical approach, improving the patient experience and increasing the treatment capacity.

C. Other Applications

Other than radiography and radiation, RL has different cancer applications, testifying to its versatility and efficacy. Gayo et al. studied RL for pre- and intra-procedure planning of prostate biopsy. They studied prostate biopsy planning with RL and imitation learning. Imitation learning is an alternative to RL for sequential decision making problems that trains a policy to mimic an expert. The results showed that the RL framework is more robust to intra-operative changes such as organ deformation. This capability makes RL a better choice for intra-procedure planning than imitation learning. RL adjusted the biopsy needle position according to dynamic conditions, enhancing the accuracy of prostate cancer treatment planning.

A cancer chemotherapeutic control approach is proposed by Alsaadi et al. based on a fuzzy-RL-based SARSA algorithm. SARSA uses the action from the current policy, and hence is more dependent on the policy than Q-learning. Their mathematical fractional order model has improved chemotherapeutic cancer control when compared with Q-learning. This discovery paves the way for more effective and less toxic personalized chemotherapy treatments.

Finally, RL could transform the game of cancer drug discovery. Tan et al. proposed the integration of reinforcement learning with biophysics, quantum computing, and other machine learning methods to solve systems pharmacology-oriented problems of personalized drug design. This combination will address the generalizability, transferability, and multi-objective optimization issues in drug development and will pave the way for more effective and tailored cancer treatments.

4. DISCUSSION

RL is a robust oncology framework for cancer screening, therapy planning, etc. Most of the RL approaches to cancer use classical RL methods in a reduced environment. There are a number of limitations and flaws that must be overcome and the newest improvements used to maximize the promise of RL for cancer. This section covers field challenges and opportunities.

A. Current Challenges

In oncology, it is difficult to define a reward function for RL. This is because the reward functions may not reflect the clinical planning goals correctly. These reward functions must reflect the diverse goals of cancer treatment. It also requires knowledge from other fields that clinical settings may not have. Delayed or sparse input complicates training of RL agents further. In oncology jobs, treatment selection feedback may take longer to arrive making it difficult to optimize performance in a single episode. Therefore, it is useful to improve the evaluation criteria and to include the judgement and opinion of

expert clinicians.

For most of the papers reviewed here, learning from a large number of trials and errors requires a large amount of computational resources and long training times. The large volumes of data necessary to train these algorithms may be hard to obtain within medical contexts due to privacy concerns, lack of data, and ethical questions regarding the use of patient data. Moreover, these models are still battling clinical validation and efficacy. Creating large and diverse datasets is hard, but necessary for robustness and accuracy.

Main issues are poor generalization, repeatability and stability. Differences in data and model architecture can cause models to behave differently and not converge. The designer experience also matters for the RL framework hyper-parameter selection. There is no publicly available source code, so it is difficult to replicate. Spinning Up in DRL is making progress. Interpretable models are important as black-box methods may hinder clinical decision-making.

The framework can deal with dynamic environments, whereas the current RL methods assume that the environment is static. However, the problem of dynamic multi-objective optimization is still open. RL needs to optimize multiple, often conflicting, dynamic rewards in a non-stationary environment. Patient responses and new medical knowledge necessitate the dynamic optimization of oncology treatment techniques. For precision medicine to tailor therapies to individual patients, scalable, generalizable models are required.

Also consider cost effectiveness. Infrastructure and training investments must be considerable to use RL models in clinical practice, and improvements in patient outcomes must justify them. This includes direct technology and staff costs and projected savings from better treatments and shorter hospital stays.

Finally, clinical acceptance demands convergence of interpretable and training models. Open decision making is required for clinicians to trust and understand model advice. Incorporation of RL into the routine of cancer practice needs to be improved in terms of training stability and explainability.

To overcome these challenges, AI researchers will need to collaborate with cancer experts to improve the robustness of algorithms and the efficiency of computing. In order to fully leverage the effect of RL on the outcomes of cancer treatment, the evaluation criteria need to be improved, the expert clinical judgment should be integrated into the RL frameworks, and innovative solutions should be developed to address the specific challenges of oncological care.

B. Future Perspectives

The future for RL in cancer looks bright with many new methods and ideas. Hierarchical RL may revolutionize cancer therapy strategies (e.g., Shen et al.) It is useful for RL agents to decompose difficult tasks into easier ones to learn and optimize. Multitasking and transfer learning can also accelerate learning and enhance effectiveness. These techniques allow models to leverage information from related tasks to improve performance on unseen tasks. The agent does not need to learn the policy from beginning, reducing training time and increasing framework generalizability.

Another possible research direction is multi-agent RL. Multiple RL agents collaborate or compete within the same environment, which makes it easier to replicate complex therapeutic relationships. Medical specialists and systems agents can work together to optimize the patient's treatment plan. Multi-agent RL may be harder if agents interact with various surrounding environments. In addition to these advances, teacher-student designs have improved learning and performance of RL agents. In cancer, a teacher model trained on rich clinical data can allow the student model to quickly adapt to new patient data and situations and improve planning of treatment and decision-making.

Future RL frameworks should include patient-centered care. In shared medical decisionmaking, the doctor and the patient jointly develop reward tables, making the incentive functions more clear and transparent, more in line with the patient's preferences and needs, and increasing the acceptability of AI tools in clinical practice. RL-based decision support with human preferences shows promise in skin cancer detection. When applied to other situations, these ideas may lead to more focused and successful actions. Active RL allows the agent to actively interact with its environment to get more information to make better decisions. This technique could improve personalized therapy, as the RL system learns from patient data and adapts its methods in real time.

In summary, enhanced RL and patient-centered approaches might improve cancer therapy. Hierarchical, active, multi-task and multi-agent RL opens new R&D opportunities. Future studies should assess patient and provider satisfaction to fully elucidate the potential for RL in cancer. These developments and constant collaboration of AI researchers with clinical experts will allow RL to

make progress in cancer therapy.

5. CONCLUSION

This study provides a comprehensive evaluation of RL in cancer. It has been useful for cancer screening and planning of therapy. RL differs from supervised and unsupervised learning by enabling adaptive decision making, personalized treatment methods, and continuous learning from patient interactions. The work dealt with reward function determination and data scarcity mitigation. Despite the difficulties, RL has a promising future. Future research needs to embed patient-centered care into RL frameworks to realize RL's transformational promise in cancer.

REFERENCES

[1]. Uc-Cetina, N. Navarro-Guerrero, A. Martin-Gonzalez, C. Weber, and S. Wermter, "Survey on reinforcement learning for language processing," Artificial Intelligence Review, vol. 56, no. 2, pp. 1543–1575, 2023.

[2]. 1. Dr. Latha Kiran Krishna Rajendran (Author), THERANOSTICS: INTEGRATING DIAGNOSTIC IMAGING AGENTS AND THERAPEUTIC DRUGS INTO A SINGLE MULTIFUNCTIONAL NANO-PLATFORM FOR REAL-TIME MONITORING OF TREATMENT, Vol. 53 No. 2 (2025): April-June 2025, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/305, DOI: https://doi.org/10.46121/pspc.53.2.31

[3]. Rajendran, O. K. (2025). Digital twin frameworks for personalized cancer progression modeling using Longitudinal data. Power System Protection and Control, 53(4), 486–501. https://doi.org/10.46121/pspc.53.4.33

[4]. N. Le, V. S. Rathour, K. Yamazaki, K. Luu, and M. Savvides, "Deep reinforcement learning in computer vision: A comprehensive survey," Artificial Intelligence Review, pp. 1–87, 2022

[5]. Hemanth Kumar, R. M. (2026). Integrated transcriptomic and machine learning framework identifies A blood-based biomarker signature for anthracycline-induced cardiotoxicity in juvenile cancer Survivors. International Journal of Drug Delivery Technology, 16(40s), 219–230. https://doi.org/10.25258/ijddt.16.40s.24

[6]. Rajendran, O. K. (2025). Deep learning for cross-modality mapping between histopathology and radiological imaging. Power System Protection and Control, 53(3), 313–328.https://doi.org/10.46121/pspc.53.3.21

[7]. G. Dulac-Arnold, N. Levine, D. J. Mankowitz, J. Li, C. Paduraru, S. Gowal, and T. Hester, "Challenges of real-world reinforcement learning: Definitions, benchmarks and analysis," Machine Learning, vol. 110, no. 9, pp. 2419–2468, 2021.

[8]. Dr. Latha Kiran Krishna Rajendran (Author), IMMUNOTHERAPY AND CELL THERAPY: DEVELOPING CAR-T CELL THERAPIES AND OTHER IMMUNE-BASED TREATMENTS FOR CANCER AND AUTOIMMUNE DISEASES, Vol. 51 No. 2 (2023): April-June 2023, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/304, DOI: https://doi.org/10.46121/pspc.51.2.7

[9]. S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," Proceedings of the IEEE, vol. 109, no. 5, pp. 820–838, 2021.

[10]. Dr. Latha Kiran Krishna Rajendran (Author), STRICT LIABILITY OR FAULT-BASED REGIMES FOR AI-CAUSED HARM? A DOCTRINAL ANALYSIS ACROSS COMMON LAW AND CIVIL LAW SYSTEMS, Vol. 52 No. 4 (2024): October-December 2024, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/312, DOI: https://doi.org/10.46121/pspc.52.4.13

[11]. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: a cancer journal for clinicians, vol. 71, no. 3, pp. 209–249, 2021.

[12]. Dr. Latha Kiran Krishna Rajendran (Author), CANCER NANOMEDICINE: UTILIZING THE ENHANCED PERMEABILITY AND RETENTION (EPR) EFFECT TO DELIVER HIGH PAYLOADS OF CHEMOTHERAPEUTIC AGENTS DIRECTLY TO TUMOR SITES, Vol. 52 No. 2 (2024): April-June 2024, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/311 , DOI: https://doi.org/10.46121/pspc.52.2.12

[13]. Dr. Latha Kiran Krishna Rajendran (Author), MECHANISMS DRIVING IMMUNOTHERAPY RESISTANCE IN COLORECTAL CANCER LIVER METASTASES, Vol. 52 No. 1 (2024): January-March 2024, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/303 DOI: https://doi.org/10.46121/pspc.52.1.5

[14]. Rajendran, O. K. (2023). Federated radiology AI models for multi-institutional cancer diagnosis without data sharing. Power System Protection and Control, 51(4), 38–54. https://doi.org/10.46121/pspc.51.4.5

[15]. Rajendran, O. K. (2023). AI-based radiogenomic models for predicting immunotherapy response in Solid tumors. Power System Protection and Control, 51(4), 24–37. https://doi.org/10.46121/pspc.51.4.4

[16]. Rajendran, L. K. K. (2026). Integrative pharmacogenomic analysis of drug response heterogeneity across cancer cell lines: Insights from large-scale GDSC data. Scientific Culture, 12(4), 7537–7546. https://doi.org/10.5281/zenodo.12426762

[17]. R. Van De Schoot, J. De Bruin, R. Schram, P. Zahedi, J. De Boer, F. Weijdema, B. Kramer, M. Huijts, M. Hoogerwerf, G. Ferdinands, et al., "An open source machine learning framework for efficient and transparent systematic reviews," Nature machine intelligence, vol. 3, no. 2, pp. 125–133, 2021.

[18]. Rajendran, O. K. (2024). Foundation model–driven precision oncology: Integrating multi-omics, radiology, and clinical data for predictive cancer care. Power System Protection and Control, 52(2),154–163. https://doi.org/10.46121/pspc.52.2.14

[19]. Rajendran, O. K. (2024). Self-supervised multimodal learning for early cancer detection across Imaging and genomics. Power System Protection and Control, 52(4), 167–178. https://doi.org/10.46121/pspc.52.4.14

[20]. Rajendran, L. K. K. (2026). Evaluating the association of cancer-related risk factors with multisystem health: Insights into fertility, cardiovascular, and renal indicators. Scientific Culture, 12(4), 7520–7527. https://doi.org/10.5281/zenodo.12426760

[21]. Rajendran, L. K. K. (2026). Impact of treatment modalities on fertility, sexual function, and Psychological outcomes in testicular cancer survivors: A comprehensive review. International Journal of Drug Delivery Technology, 16(30s), 447–453. https://doi.org/10.25258/ijddt.16.30s.43

[22]. Rajendran, L. K. K. (2026). From prediction to practice: A machine learning–based clinical decision Support tool for bevacizumab risk stratification in oncology. International Journal of Drug Delivery Technology, 16(30s), 414–429. https://doi.org/10.25258/ijddt.16.30s.40

[23]. Rajendran, L. K. K. (2026). From prediction to precision: An externally validated deep learning–based Survival and adjuvant therapy recommendation system for resected stage III non-small cell lung Cancer. International Journal of Drug Delivery Technology, 16(30), 430–438. https://doi.org/10.25258/ijddt.16.30.41

[24]. Rajendran, L. K. K. (2026). Interpretable machine learning for early mortality prediction in acute Myeloid leukemia: A decision tree–based retrospective cohort study. International Journal of Drug Delivery Technology, 16(40s), 231–241. https://doi.org/10.25258/ijddt.16.40s.25

[25]. 12. Rajendran, L. K. K. (2026). Machine learning–driven symptom-based cancer risk stratification: A Systematic review of clinical prediction models and methodological rigor. International Journal of Drug Delivery Technology, 16(40s), 242–253. https://doi.org/10.25258/ijddt.16.40s.26

[26]. S. K. Zhou, H. N. Le, K. Luu, H. V. Nguyen, and N. Ayache, "Deep reinforcement learning in medical imaging: A literature review," Medical image analysis, vol. 73, p. 102 193, 2021.

[27]. T. M. Moerland, J. Broekens, A. Plaat, C. M. Jonker, et al., "Modelbased reinforcement learning: A survey," Foundations and Trends® in Machine Learning, vol. 16, no. 1, pp. 1–118, 2023.

[28]. J. Wang, J. Miao, X. Yang, R. Li, G. Zhou, Y. Huang, Z. Lin, W. Xue, X. Jia, J. Zhou, et al., "Auto-weighting for breast cancer classification in multimodal ultrasound," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23, Springer, 2020, pp. 190–199.

[29]. Yala, P. G. Mikhael, C. Lehman, G. Lin, F. Strand, Y.-L. Wan, K. Hughes, S. Satuluru, T. Kim, and I. Banerjee, "Optimizing riskbased breast cancer screening policies with reinforcement learning," Nature medicine, vol. 28, no. 1, pp. 136–143, 2022.

[30]. Kolla and P Venugopal, "Breast cancer diagnosis through knowledge distillation of swin transformer-based teacher–student models," Machine Learning: Science and Technology, vol. 4, no. 4, p. 045 047, 2023.

[31]. Barata, V. Rotemberg, N. C. Codella, P. Tschandl, C. Rinner, B. N. Akay, Z. Apalla, G. Argenziano, A. Halpern, and A. Lallas, "A reinforcement learning model for ai-based decision support in skin cancer," Nature medicine, vol. 29, no. 8, pp. 1941–1946, 2023.

[32]. Xu, Y. Song, D. Zhang, L. K. Bittencourt, S. H. Tirumani, and S. Li, "Spatiotemporal knowledge teacher–student reinforcement learning to detect liver tumors without contrast agents," Medical Image Analysis, vol. 90, p. 102 980, 2023

[33]. Shen, D. Nguyen, L. Chen, Y. Gonzalez, R. McBeth, N. Qin, S. B. Jiang, and X. Jia, "Operating a treatment planning system using a deep-reinforcement learning-based virtual treatment planner for prostate cancer intensity-modulated radiation therapy treatment planning," Medical physics, vol. 47, no. 6, pp. 2329–2336, 2020.

[34]. Shen, L. Chen, and X. Jia, "A hierarchical deep reinforcement learning framework for intelligent automatic treatment planning of prostate cancer intensity modulated radiation therapy," Physics in Medicine & Biology, vol. 66, no. 13, p. 134 002, 2021.

[35]. Niraula, J. Jamaluddin, M. M. Matuszak, R. K. T. Haken, and I. E. Naqa, "Quantum deep reinforcement learning for clinical decision support in oncology: Application to adaptive radiotherapy," Scientific reports, vol. 11, no. 1, p. 23 545, 2021.

[36]. P. Kafaei, Q. Cappart, M.-A. Renaud, N. Chapados, and L.-M. Rousseau, "Graph neural networks and deep reinforcement learning for simultaneous beam orientation and trajectory optimization of cyberknife," Physics in Medicine & Biology, vol. 66, no. 21, p. 215 002, 2021.

[37]. J. Gayo, S. U. Saeed, E. Bonmati, D. C. Barratt, M. J. Clarkson, and Y. Hu, "The distinct roles of reinforcement learning between pre-procedure and intra-procedure planning for prostate biopsy," International Journal of Computer Assisted Radiology and Surgery, pp. 1–10, 2024.

[38]. E. Alsaadi, A. Yasami, C. Volos, S. Bekiros, and H. Jahanshahi, "A new fuzzy reinforcement learning method for effective chemotherapy," Mathematics, vol. 11, no. 2, p. 477, 2023.

[39]. Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," IEEE transactions on neural networks and learning systems, vol. 32, no. 1, pp. 4–24, 2020.

[40]. R. K. Tan, Y. Liu, and L. Xie, "Reinforcement learning for systems pharmacology-oriented and personalized drug design," Expert opinion on drug discovery, vol. 17, no. 8, pp. 849–863, 2022.