Deep reinforcement learning for diseases detection : A literature review

Hadjer BOUREKOUCHE

LIST Laboratory, Department of Electrical Systems Engineering, Faculty of Technology, University M'Hamed Bougara of Boumerdes, Algeria

Email : h.bourekouche@univ-boumerdes.dz

Abstract **-** The use of artificial intelligence (AI) in the evaluation of medical images has resulted in accurate assessments being automatically conducted. This has reduced the workload of physicians, decreased diagnostic errors and turnaround times, and improved the performance in the prediction and detection of various diseases. Deep reinforcement learning (DRL), which is a new type of AI, has achieved a dramatic increase in the number of applications, including medical imaging tasks. Numerous DRL diagnostic algorithms have been developed for detecting various diseases such as lung cancer, heart disease, Alzheimer's disease, dengue disease, and Parkinson's disease. This study focuses on the recent developments in deep reinforcement learning for the detection and diagnosis of various diseases.

Keywords **-** Artificial intelligence, deep reinforcement learning, DRL, disease detection, medical imaging.

I. INTRODUCTION

Owing to the recent integration of deep learning models with reinforcement learning algorithms in various applications, deep reinforcement learning has achieved considerable success [1]. To address the perception and decision-making issues of complex systems, deep learning is combined with reinforcement learning [2]. The rapid development of DRL has led to notable performance improvements in various fields, such as gaming, robotics, natural language processing, and computer vision [3], particularly in medicine, for disease detection.

Medical item detection and segmentation are essential pre-processing procedures in the clinical workflow for diagnosis and therapy planning [4]. Although deep learning methods have shown impressive success in this area, they have several drawbacks, including computational restrictions, inadequate parameter tuning, and poor generalization. The newest artificial intelligence technique, deep reinforcement learning, has the ability to significantly improve upon the shortcomings of conventional methods of detection and segmentation while also producing accurate findings [5].

In the subject of disease detection, this study provides a review of the current developments in deep reinforcement learning. Initially, several fundamental ideas regarding deep learning, reinforcement learning, and deep reinforcement learning are presented in Section II. In Section III, several DRL algorithms, including DQN, DDPG, PPO, A2C, and A3C, the most recent DRL model-free and model-based algorithms are presented then compared in Section IV. This is followed by the most recent DRL medical imaging tasks in Section V, which is the main section of the literature review, followed by a discussion in Section V. Finally, the underlying issues and future prospects of DRL are examined.

II. PRELIMINARIES

In this section, we introduce the basic elements of machine learning, deep learning, and deep reinforcement learning.

A) Machine learning ML

Machine learning is a fascinating field in the fields of computer science and engineering. Because it makes it possible to extract useful patterns from examples, which is a feature of human intelligence, it is regarded as a subset of

artificial intelligence. The fundamental goal of ML methods is to create algorithms that can learn from and make predictions based on data [6]. Supervised, unsupervised, and reinforcement learning are the three main categories of machine learning, each of which is further divided into subcategories [7], as illustrated in Figure. 1.

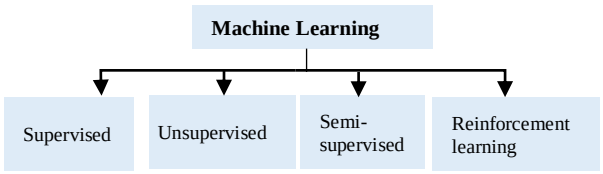

Fig. 1. ML Classification.

The following is a brief discussion of each algorithm :

✓ *Supervised Learning*

Through supervised learning, the machine is trained using examples. The controller provides predefined data, including the desired outcomes, to the machine learning algorithm. The algorithm must devise a route to the inputs and outputs after the dataset is discovered [8].

✓ *Semi-supervised Learning*

Semi-supervised learning is similar to supervised learning, except that it uses datasets that have been marked and those that have not. Unlabeled data lack the relevant tags that labeled data have, but labeled data have them so that the machine-learning algorithm can understand it. Using this method, machine learning systems can learn to recognize unlabeled data [8, 9].

✓ *Unsupervised Learning*

The machine learning program examines the data provided for patterns or correlations. No aid was provided. The machine examined readily available data to uncover linkages and connections. Large datasets are left up to the machine-learning algorithm to comprehend and react to in an unsupervised learning process [8].

✓ *Reinforcement Learning*

Reinforcement Learning is an area of machine learning concerned with sequential decision problems. Specifically, a learner or agent interacts with an environment by taking action (Fig. 2), and the aim of the agent is to maximize its expected cumulative reward [10]. The long-term benefits are maximized by reinforcement learning algorithms. The definition of a policy may or may not include the action plan of an agent [11].

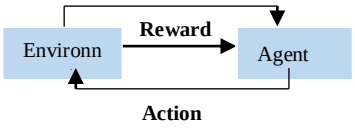

Fig. 2. Structure of RL.

Mathematical formulation of Reinforcement Learning Problems :

The process for modeling issues in reinforcement learning is called the Markov Decision Process (MDP) [12], and is described in Reinforcement Learning. This is used to mathematically represent the sequential decision-making issues. A collection of states creates an environment for reinforcement learning. Markov decision processes (MDPs) are a formalization of sequential decision processes, where an agent in a state (s) performs an action (a), which may cause it to shift states (s'), and then gets a reward (r). It must make a fresh judgment on what action to do in the new state until a termination requirement is satisfied.

An MDP is represented by a tuple $M = (S, A, \phi, R)$, where[13] :

S is a finite set of states ($s_i \in S, i = \{1, \ldots, n\}$),

A is a finite set of actions, which can depend on each state ($a_j\ (s_i) \in A, j = \{1, \ldots, m\}$),

$R(s, a)$ is a reward function that describes the state and maps each state-action pair to a number (reward) reflects the desirability of the state $(R : S \times A \rightarrow R)$[13].

$\phi\ (s'|s, a)$ is a state transition function that gives us the probability of reaching state $s' \in S$ when taking action $a \in A$ in state $s \in$ S (: S × A → S)

Additional elements include: Policy (π): outlines the agent's behavior. It is a mapping from states to actions that might potentially be stochastic($\pi(s) \rightarrow a$) which dictates which action to take on each state[13].

Value function (V_π /Q_π): Through equations (1) and (2), it is the expected accumulated rewards that an agent anticipates to receive in a state[13] $s\ (V_\pi(s))$.

$$V_\pi(s) = \mathbb{E}_\pi[G_t|S_t = s] = \mathbb{E}_\pi[\sum_{k=0}^{\infty} \gamma^k\, r_{t+k+1}|S_t] \quad (1)$$

Or in a state s taking an action a($Q_\pi(s,a)$):

$$Q_\pi(s,a)=\mathbb{E}_\pi[G_t|S_t=s,a_t=a] = \mathbb{E}_\pi[\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}|s_t=s,a_t=a] \quad (2)$$

and following the policy π onward, where γ is a discount factor and G_t is the total discounted reward.

B) Deep Learning

Deep learning is a subset of ML approaches that use artificial neural networks (ANNs) that are inspired by the structure of neurons in the human brain [6]. The term "deep" originally meant that an artificial neural network had many layers; however, this meaning of deep has evolved over time. Convolutional Neural Networks (CNNs), Long Short-Term Memory Networks (LSTMs), Recurrent Neural Networks (RNNs), and autoencoders are among the most well-known deep neural networks [2].

CNNs have generally shown remarkable advantages in many learning tasks, such as those involving visual and audio inputs [6].

C) Deep Reinforcement Learning

Reinforcement learning and deep learning are the two components of DRL. According to the aforementioned sections, the deep reinforcement learning method can make the best judgments but lacks perception capacity, whereas the reinforcement learning method provides a robust representation of the environment but weak decision-making skills [2]. Finding a technique to depict the environment and simultaneously choose the optimal course of action is crucial.

Deep reinforcement learning combines the perception ability of deep learning with the decision-making ability of reinforcement learning [14].

Figure 3 depicts a simple deep reinforcement-learning architecture. The agent interacts with the environment and observes the state s_t at each time step t, where $s_t \in S$ based on the deep learning method. Consider each action's reward function before choosing action (a) from a list of possible actions $A = (1, 2, \ldots, k)$ when a is at state $s_t \in S$, where S is the set of states. As a result of the action, the agent will receive a reward $rt \in R$. The environment responds to this action and the agent observes the following state: $s_{t+1} \in S[2]$.

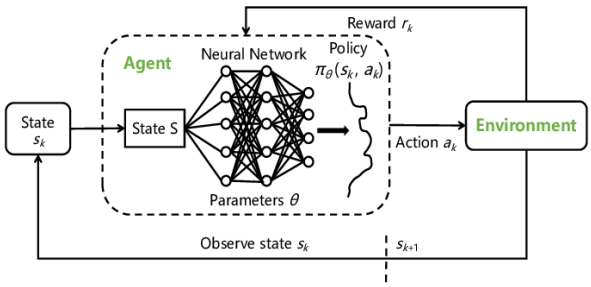

Fig. 3. DRL Framework.

III. DRL ALGORITHMS

As shown in Fig. 4, deep reinforcement learning has been used to estimate value functions, policy functions, or both (actor-critic technique) [13]. Policy gradient methods, as opposed to value-based methods that do not attempt to explicitly optimize across a policy space, may learn parameterized policies without intermediary value estimation. Optimal value reinforcement learning algorithms include Q learning, DQN, double DQN, dueling DQN. Policy gradient algorithms include the policy gradient method, actor-critic algorithm, A3C, A2C, DPG, DDPG [15], TRPO, and PPO.

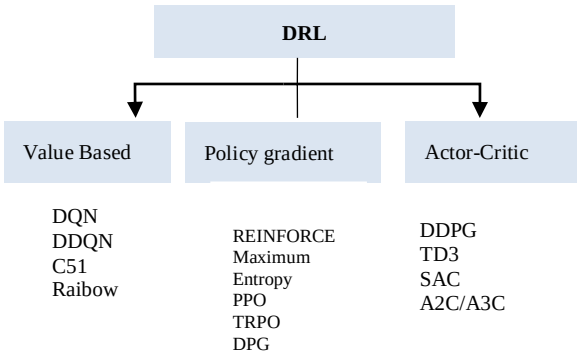

Fig. 4. An overview of some deep reinforcement learning algorithms.

A) Value Based Methods

Whereas policy-based algorithms directly update the control policies for an agent without employing value estimates, value-based algorithms develop control policies for an agent by predicting the likelihood of the agent reaching a certain state [16]. Value-based methods are reviewed briefly in this section.

1) Deep Q-Network (DQN)

The Deep Q-network (DQN), developed by Minh et al.[17], is the first deep reinforcement learning model that has been successfully trained in practice[18]. It is a multi-layered neural network that produces a vector of action values for a given input state and is built by fusing a convolutional neural network with three convolutional layers, two fully connected layers, and a Q-learning algorithm.

A deep Q network is composed of three parts: a Q network that determines the policy, a target Q network that generates target Q values for the loss function, and a replay memory that stores training samples.

2) Double DQN

A modified form of the DQN and tabular double Q-learning techniques is called a double DQN [19]. To facilitate exposition, let y^{DQN} denote the target Q-value [19], y^{DQN} is given by equation (3) as:

$$y^{DQN} = r + \gamma \max_{a' \in A} Q\,(s', a'; \phi') \quad (3)$$

The same network with parameter ϕ' is used to choose (greedy) actions and evaluate actions with this goal Q-value, this might result in the selection of overly optimistic policies, which, in turn, produces overly optimistic value estimations. By employing the target network as a second-value function approximator, the double DQN seeks a divorce action assessment from the action selection. Consequently, the anticipated target Q-value was estimated as follows :

$$y^{DoubleDQN} = r + \gamma Q(s', \operatorname*{argmax}_{a' \in A} Q(s', a'; \phi'); \phi') \quad (4)$$

3) Dueling DQN

According to the dueling network design, which consists of two sequences of completely linked convolution layers, Dueling DQN is another modified version of DQN [19]. The state value and advantage functions were estimated individually using these two sequences. The advantage function $A_\pi(s, a)$, which is for an action a, $\pi(s)$ from a state s based on policy π, is the difference between the Q-value associated with this state-action pair $Q_\pi(s, a)$ and the state value function $V_\pi(s)$. The Q function is calculated by combining the estimations of the state value function and the advantage function from two different sequences [19].

$$Q(s, a; \phi, \alpha, \beta) = V(s; \phi, \beta) + (A(s, a; \phi, \beta) - \frac{1}{|A|} \sum_{a'} A(s, a'; \phi, \alpha)) \quad (5)$$

where α and β are the parameters of two fully connected layers. The key concept underlying the dueling network value architecture is preventing the needless estimation of the value of each action option from a state.

B) Policy Gradient Methods

This section describes policy gradient (PG) techniques, a type of policy-based algorithm that is extensively employed in reinforcement learning.

1) Stochastic Policy Gradient SPG

The most well-known class of continuous action reinforcement learning algorithms is the policy-gradient algorithm. The fundamental principle of these algorithms is to modify policy parameters θ in the direction of the performance gradient $\nabla_\theta J(\pi_\theta)$. The fundamental result underlying these algorithms is the policy gradient theorem :

$$\nabla_\theta J(\pi_\theta) = \int_s \rho^\pi(s) \int_A \nabla_\theta \pi_\theta(a|s) Q^\pi(s, a) da ds \quad (6)$$

Surprisingly, the policy gradient is straightforward. In particular, the policy gradient does not rely on the gradient of the state distribution $\rho^\pi(s)$, even when the state distribution relies on the policy parameters.

2) Deterministic Policy Gradient (DPG)

The predicted gradient of the action-value function is based on the DPG approach. The gradient of the deterministic policy can be estimated without employing an integral term over action space. It has been shown that, in a high-dimensional action space, DPG algorithms can outperform SPG algorithms.

3) Proximal Policy Optimization (PPO)

The Trust Region Optimization method is used in Proximal Policy Optimization (PPO)[20], a cutting-edge policy gradient method, to ensure that each step in the policy optimization process leads to a better or the same policy, but not a worse one. Importance sampling is used to introduce the sample inefficiency problem.

Among the actor-critic DRL algorithms, it has been shown to be a particularly stable and dependable method. When computing gradients, the PPO algorithm actively restricts the modifications to the policy. The policy $\pi(A_t|s_t,\theta)$ maps system states (s_t) to Actions (A_t) according to the current parameters θ of a neural net (the "actor"), which, together with the parameters of another neural net (the "critic") ω, are updated during the training process.

C) Actor-Critic methods

The actor changes the policy distribution using policy gradients in the actor's critic architecture, and the critic calculates the value function for the current policy[21].

1) Deep deterministic policy gradient (DDPG)

A model-free DRL algorithm known as a DDPG may work in continuous action space and is regarded as an actor-critic based algorithm. The actor acts in an environment to optimize and approximate its intrinsic policy using the value function supplied by the critic based on the deterministic policy gradient algorithm. After receiving rewards from the surrounding environment, the critic estimates the value function, which is then used to optimize the actor's policy approximation. Using a mini-batch of the same, DD will be able to make the most of its time in the field.

2) Advantage actor-critic (A2C)

The on-policy family of policy gradient algorithms includes A2C. In other words, we cannot learn the value function by adhering to a different policy, indicating that we learn the value function for one policy while adhering to it. If we were to employ experience replay, for instance, we would adopt a different policy because, by learning from old data, we would be utilizing data produced by a policy (the network) that was marginally different from the state at hand.

3) Advantage actor-critic (A3C)

A3C is an asynchronous version of the advantage actor-critic (A2C). The policy and value functions are updated after every t_{max} actions or when a terminal state is reached. The update performed by the algorithm can be seen as:

$$\nabla_{\theta'} log\pi\left(a_t|s_t;\theta'\right)A(s_t,a_t;\theta,\theta_v) \quad (7)$$

where $A(s_t,a_t;\theta,\theta_v)$ is an estimate of the advantage function given by :

$$\sum_{i=0}^{k-1}\gamma^i r_{t+i}+\gamma^k V(s_{t+k};\theta_v) \quad (8)$$

where k varies from state to state and is upper-bounded by t_{max} [17].

IV. COMPARISON

In this section, we compare the various deep reinforcement learning algorithms discussed in Section III. We discuss the main characteristics, performance of the convergence speed, and advantages and drawbacks of different DRL. Table I summarizes the results of this comparison. In general, it can be observed that :

- PPO is more sensitive to modifications than other approaches but offers a higher rate of convergence and performance. Because the DQN alone is unstable and provides poor convergence, it requires several additions.

- A3C is highly helpful when there is a requirement to teach the notion of transfer learning in comparable settings and considerable computing power is available.

V. RELATED WORKS ON APPLICATIONS OF DRL IN MEDICINE : DISEASES DETECTION

Robotics, computer games, and even medicine have benefited greatly from deep reinforcement learning. Medical image processing and analysis include complicated and crucial procedures, such as medical object detection and segmentation, which are examples of DRL applications in the medical industry. Their goals are to identify regions in medical images that have particular special meanings (such as lesions and vertebral bodies) and to classify the pixels in those regions into desirable objects, thereby providing a solid foundation for clinical diagnosis and pathological research to help doctors arrive at a more accurate diagnosis. This section focuses on the use of deep reinforcement learning to detect diseases, and all

DRL algorithms, in general, are solely supported by numerical experiments without any theoretical assumptions. Additionally, these algorithms not only use the experience replay and target network strategies proposed in the original DQN but also create new strategies to improve performance.

DRL for Lung Cancer detection

Lung cancer is the most prevalent malignant tumor worldwide. Tumor localization is a crucial area where deep reinforcement learning can be used for the diagnosis and treatment of lung cancer. Accurate tumor localization is particularly important for lung cancer diagnosis, because it may enhance surgery and lower the risk of recurrence.

Liu et al.[18] described various common deep reinforcement learning models[18] that have the potential to be utilized for lung cancer diagnosis and detection. They focused on value-based deep reinforcement learning models that have this potential, such as Deep Q-networks and the hierarchical deep Q-network (h-DQN), which blends deep reinforcement learning with goal-driven intrinsic motivation and hierarchical action-value functions.

Ali et al.[22] created and validated a deep reinforcement learning model based on deep artificial neural networks for the early detection of lung nodules in thoracic CT scans[23], drawing inspiration from the AlphaGo system. The LIDC/IDRI database, which was maintained using lung nodule analysis (LUNA), served as the training data for our model.

- ✓ *DRL for breast lesions detection*

Breast cancer is one of the malignancies with the highest number of recent diagnoses worldwide. A detection model based on a DQN trained with reinforcement learning is presented by Maicas et al.[24] and is capable of accelerating the inference time of lesion detection from the breast[24]. This may shorten the inference time while maintaining the breast DCE-MRI cutting-edge accuracy for lesion detection. This research was based on the method of Caicedo and Lazebnik et al. [25].

- ✓ *DRL for Brain tumor detection*

Using the BraTS brain tumor imaging database[26], Stember et al.[27] trained a deep Q-network (DQN) on 70 post-contrast T1-weighted 2D image slices[26] to predict the brain tumor location.

- ✓ *DRL for Coronary heart disease prediction*

The prediction of coronary heart disease is a crucial and difficult issue in the medical world. The asynchronous advantage actor-critic (A3C) method [28] is utilized to effectively address the issue in Li et al. [28] proposed mixed reinforcement multitask progressive time-series network (CRMPTN) model to predict the grade of coronary heart disease using a heart color Doppler echocardiography report[29].

VI. DISCUSSION

DRL is a sequential procedure with shifting attention concentration that progressively accumulates evidence of confidence while searching for the desired object by fusing the potent decision-making abilities of RL with the potent perception of DL. Consequently, DRL may concurrently use its search technique and the presence of an object of interest to identify where the desired object is related to tasks involving the detection and segmentation of medical objects using a sequential procedure. Detection and segmentation methods may concentrate on a particular region and provide a more accurate result by using the object location as a prior. DRL successfully avoids slipping into suboptimal situations, in contrast to the conventional DL method, which directly predicts the object location using a mapping based on appearance.

In general, all DRL algorithms are supported solely by numerical experiments without any theoretical assumptions. Additionally, these algorithms not only use the experience replay and target network strategies proposed in the original DQN but also create new strategies to improve performance.

Table 1. DRL ALGORITHMS COMPAARISON

Algorithm	DQN	A2C	A3C	PPO	DDPG
Description	Deep Q netw	Advantage actor-critic	Advantage actor-critic	Proximal Policy Optimization	Deep deterministic policy gradient
Value Based/Policy gradient/Actor-Critic	Value based	Actor critic	Actor critic	Policy gradient	Actor critic
Policy	Off-policy	Off-policy	On-Policy	On-policy	Off-Policy
Action Space	Discrete	Discrete/Continus	Discrete/Continus	Discrete/Continus	Continus
Speed of convergence	Reasonable	Good	Good	Excellent	Good
Stability of convergence	Reasonable	Reasonable	Reasonable	Good	Good
Pros	- Uses memory replay –sample efficiency - Allow any exploration strategy - Requires a Target network for stability - High efficiency - Solves problems with high-dimensional observation spaces (Timothy,2019)	Can be used for environments with either discrete or continuous action spaces	- Provides faster multi-processing - The convergence rate is faster due to multiple instances running concurrently. - Possible use of RNNs - Can be used for environments with either discrete or continuous action spaces	- Simpler to implement. - Can be used for environments with either discrete or continuous action spaces - Can be used for environments with either discrete or continuous action spaces	Replay buffer; Actor-critic architecture; Deterministic policy; Unbiased critic network; Exploration noise
Cons	- One step Learning: Slow to propagate information - Slow convergence - Accumulate error -Requires Prioritized Experience Replay for efficient sampling - It can only handle discrete and low-dimensional action spaces - No use of RNNS	- Do not use memory replay - No exploration			- Can only be used for environments with continuous action spaces

VII. CONCLUSION

Particularly for challenging tasks involving ongoing decision-making, deep reinforcement learning models may generally outperform reinforcement learning approaches. Because these models have been shown to be effective in other areas of medical image analysis, they are expected to be employed to tackle the issue of disease detection and provide positive outcomes. These DRL-based methods have significantly increased the accuracy of the results, and several of them are now considered industry standards. In this article, we provide a comprehensive summary of current research on deep reinforcement learning, with an emphasis on its usage in medical imaging, particularly for the detection of diseases. As a result, we gained knowledge of the distinctive features of each DRL-based technique, including policy gradients, value-based methods, and actor-critic methods.

VIII. REFERENCES

[1] Z. Liu, C. Yao, H. Yu, and T. Wu, “Deep reinforcement learning with its application for lung cancer detection in medical Internet of Things,” Future Gener. Comput. Syst., vol. 97, pp. 1–9, Aug. 2019, doi: 10.1016/j.future.2019.02.068.

[2] W. Xu, L. Chen, and H. Yang, "A Comprehensive Discussion on Deep Reinforcement Learning," in 2021 International Conference on Communications, Information System and Computer Engineering (CISCE), May 2021, pp. 697–702. doi: 10.1109/CISCE52179.2021.9445914.

[3] S. K. Zhou, H. N. Le, K. Luu, H. V. Nguyen, and N. Ayache, "Deep reinforcement learning in medical imaging: A literature review." arXiv, Mar. 05, 2021. doi: 10.48550/arXiv.2103.05115.

[4] H. Bourekouche, S. BELKACEM, and N. Messaoudi, "Lightweight Medical Image Encrypting and Decrypting Algorithm Based on the 3D Intertwining Logistic Map," Int. J. Inform. Appl. Math., vol. 6, Jan. 2024, doi: 10.53508/ijiam.1405959.

[5] D. Zhang, "Deep Reinforcement Learning in Medical Object Detection and Segmentation," Electron. Thesis Diss. Repos., Nov. 2020, [Online]. Available: https://ir.lib.uwo.ca/etd/7423

[6] Z. Gao and X. Wang, "Deep Learning," in EEG Signal Processing and Feature Extraction, L. Hu and Z. Zhang, Eds., Singapore: Springer, 2019, pp. 325–333. doi: 10.1007/978-981-13-9113-2_16.

[7] J. Jagannath, A. Jagannath, S. Furman, and T. Gwin, "Deep Learning and Reinforcement Learning for Autonomous Unmanned Aerial Systems: Roadmap for Theory to Deployment." arXiv, Sep. 08, 2020. doi: 10.48550/arXiv.2009.03349.

[8] "Machine Learning vs. Deep Learning: The Ultimate Comparison." Accessed: Jul. 11, 2023. [Online]. Available: https://www.iteratorshq.com/blog/machine-learning-vs-deep-learning-the-ultimate-comparison/

[9] P. P. R, "CSE2003_Winter18_19." Accessed: Jul. 11, 2023. [Online]. Available: https://padmapriyadsa19.blogspot.com/2019/03/discuss-about-role-of-algorithms-in.html

[10] A. Jonsson, "Deep Reinforcement Learning in Medicine," Kidney Dis., vol. 5, no. 1, pp. 18–22, Oct. 2018, doi: 10.1159/000492670.

[11] J. Mulani, S. Heda, K. Tumdi, J. Patel, H. Chhinkaniwala, and J. Patel, "Deep Reinforcement Learning Based Personalized Health Recommendations," in Deep Learning Techniques for Biomedical and Health Informatics, S. Dash, B. R. Acharya, M. Mittal, A. Abraham, and A. Kelemen, Eds., in Studies in Big Data. , Cham: Springer International Publishing, 2020, pp. 231–255. doi: 10.1007/978-3-030-33966-1_12.

[12] Z. Ding, Y. Huang, H. Yuan, and H. Dong, "Introduction to Reinforcement Learning," in Deep Reinforcement Learning: Fundamentals, Research and Applications, H. Dong, Z. Ding, and S. Zhang, Eds., Singapore: Springer, 2020, pp. 47–123. doi: 10.1007/978-981-15-4095-0_2.

[13] E. F. Morales, R. Murrieta-Cid, I. Becerra, and M. A. Esquivel-Basaldua, "A survey on deep learning and deep reinforcement learning in robotics with a tutorial on deep reinforcement learning," Intell. Serv. Robot., vol. 14, no. 5, pp. 773–805, Nov. 2021, doi: 10.1007/s11370-021-00398-z.

[14] Y. Lei, X. Zhai, and B. V. D. Kumar, "Distributed System Computing Resource Scheduling Algorithm Based on Deep Reinforcement Learning," Int. J. Comput. Inf. Eng., vol. 17, no. 3, pp. 180–185, Mar. 2023.

[15] "research.pdf - I.Define the following 1.Artificial Intelligence Artificial intelligence is the simulation of human intelligence processes by | Course Hero." Accessed: Jul. 12, 2023. [Online]. Available: https://www.coursehero.com/file/159617846/research pdf/

[16] C. Chu, K. Takahashi, and M. Hashimoto, "Comparison of Deep Reinforcement Learning Algorithms in a Robot Manipulator Control Application," in 2020 International Symposium on Computer, Consumer and Control (IS3C), Nov. 2020, pp. 284–287. doi: 10.1109/IS3C50286.2020.00080.

[17] V. Mnih et al., "Asynchronous Methods for Deep Reinforcement Learning." arXiv, Jun. 16, 2016. doi: 10.48550/arXiv.1602.01783.

[18] "Deep reinforcement learning with its application for lung cancer detection in medical Internet of Things - PDF Free Download," c.coek.info. Accessed: Jul. 11, 2023. [Online]. Available: https://c.coek.info/pdf-deep-reinforcement-learning-with-its-application-for-lung-cancer-detection-in-me.html

[19] N. Parvez Farazi, B. Zou, T. Ahamed, and L. Barua, "Deep reinforcement learning in transportation research: A review," Transp. Res. Interdiscip. Perspect., vol. 11, p. 100425, Sep. 2021, doi: 10.1016/j.trip.2021.100425.

[20] M. Saeed, M. Nagdi, B. Rosman, and H. H. S. M. Ali, "Deep Reinforcement Learning for Robotic Hand Manipulation," in 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Feb. 2021, pp. 1–5. doi: 10.1109/ICCCEEE49695.2021.9429619.

[21] A. Mosavi, P. Ghamisi, Y. Faghan, and P. Duan, "Comprehensive Review of Deep Reinforcement Learning Methods and Applications in Economics," Mar. 2020. doi: 10.20944/preprints202003.0309.v1.

[22] J. Ali et al., "Lung Nodule Detection via Deep Reinforcement Learning," Front. Oncol., vol. 8, 2018, Accessed: Jul. 12, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fonc.2018.00108

[23] G. Han, Y. Kong, H. Wu, and H. Li, "Deep Reinforcement Learning Method for 3D-CT Nasopharyngeal Cancer Localization with Prior Knowledge," Appl. Sci., vol. 13, no. 14, Art. no. 14, Jan. 2023, doi: 10.3390/app13147999.

[24] G. Maicas, A. P. Bradley, J. C. Nascimento, I. Reid, and G. Carneiro, "Deep Reinforcement Learning for Detecting Breast Lesions from DCE-MRI," in Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics, L. Lu, X. Wang, G. Carneiro, and L. Yang, Eds., in Advances in Computer Vision and Pattern Recognition. , Cham: Springer International Publishing, 2019, pp. 163–178. doi: 10.1007/978-3-030-13969-8_8.

[25] J. C. Caicedo and S. Lazebnik, “Active Object Localization with Deep Reinforcement Learning.” arXiv, Nov. 18, 2015. doi: 10.48550/arXiv.1511.06015.

[26] M. Ibrahim and R. Elhafiz, “Integrated Clinical Environment Security Analysis Using Reinforcement Learning,” Bioeng. Basel Switz., vol. 9, no. 6, p. 253, Jun. 2022, doi: 10.3390/bioengineering9060253.

[27] J. Stember and H. Shalu, “Deep reinforcement learning to detect brain lesions on MRI: a proof-of-concept application of reinforcement learning to medical images.” arXiv, Aug. 06, 2020. doi: 10.48550/arXiv.2008.02708.

[28] W. Li, M. Zuo, H. Zhao, Q. Xu, and D. Chen, “Prediction of coronary heart disease based on combined reinforcement multitask progressive time-series networks,” Methods, vol. 198, pp. 96–106, Feb. 2022, doi: 10.1016/j.ymeth.2021.12.009.

[29] X. Zhang et al., “An accurate diagnosis of coronary heart disease by Catboost, with easily accessible data,” J. Phys. Conf. Ser., vol. 1955, p. 012027, Jun. 2021, doi: 10.1088/1742-6596/1955/1/012027