INSTITUTE FOR ADVANCED ECE RESEARCH
RESEARCH ARTICLE

IAECES Journal of Electronics and Communication Engineering
Vol. 1, No. 1, Jan - Dec 2026 (pp.105-112)

Reinforcement Learning-Based Adaptive Resource Management in Next-Generation Wireless Communication Networks

K. Maidanov

Department of Electrical and Computer Engineering, Ben-Gurion University, Beer Sheva, Israel

KEYWORDS:

Reinforcement Learning,
Resource Allocation,
Wireless Networks,
Deep Q-Network,
Adaptive Scheduling,
6G Communication

ARTICLE HISTORY:

Submitted : 18.01.2026
Revised : 16.02.2026
Accepted : 13.03.2026

DOI:
https://doi.org/10.17051/IJECE/01.01.14

Abstract

Intelligent and adaptive resource management mechanisms will be needed in next generation wireless communication networks in order to deal with the high user density, dynamic traffic patterns and stringent Quality of Service (QoS) specifications. Traditional optimization and other heuristic-driven methods have drawbacks in the ability to cope with changing and stochastic nature of the wireless environment and hence perform poorly under real world conditions. The present paper represents a reinforcement learning (RL)-based framework of adaptive management of network resources, such as bandwidth and transmission power, based on the dynamic allocation. The issue is modeled as a Markov Decision Process (MDP) which allows intelligent agent to acquire optimal allocation policies by continuous interaction with the network environment. An architecture based on Deep Q-Network (DQN) is adopted to estimate the optimal action-value function and facilitate real-time decision-making. The suggested framework optimally operates throughput, latency and energy efficiency in the changing network conditions. The simulation results have shown a 30 percent improvement in throughput, 35 percent decrease in latency and 25 percent increase in energy efficiency of the RL-based approach over the conventional and heuristic-based approaches. The outcomes confirm the effectiveness, scalability and strength of the proposed framework and it is a promising framework to manage adaptive resources in the next-generation wireless communication system.

Author's e-mail: maidanov.k@gmail.com

How to cite this article: Maidanov K, Reinforcement Learning-Based Adaptive Resource Management in Next-Generation Wireless Communication Networks, IAECES Journal of Electronics and Communication Engineering, Vol. 1, No. 1, 2026 (pp.105-112).

1. Introduction

This has made the process of managing resources much more complicated with the fast development of wireless communication at next-generation networks, 5G and the emerging 6G systems, which feature an ultra-dense deployment, a heterogeneous device base, and a wide range of Quality of Service (QoS) needs. It is important to efficiently distribute limited resources of these networks including spectrum, transmission power, and bandwidth to maintain high data rates, ultra-low latency, and energy efficiency [1], [2]. Yet, the time-varying channel conditions, coupled with changing traffic demand, result in dynamic and highly unpredictable wireless environments that have become very challenging to traditional resource allocation methods. Conventional methods, such as optimization-based and heuristic methods tend to assume simplified system models and worsted conditions, and are not suitable in real time adaptation in complex network situation [3], [4]. These techniques do not usually perform optimally when there is uncertainty and are unable to react well to sudden shifts in the environment. The techniques of machine learning, especially reinforcement learning (RL) in the management of wireless resources have been studied recently as they can develop optimal policies through interactions with the environment [5], [6]. Nevertheless, the current RL-based methods have a number of limitations in spite of these improvements. Most research concentrates on problem solving of single-objective optimization, including throughput maximization or power minimization, not taking into account trade-offs between various performance measures [7]. Also scalability and real time adaptability are still issues of concern, particularly in large scale and heterogeneous wireless networks [8]. Moreover, a number of works do not provide a thorough comparison

of the results with the baseline techniques, and do not reflect the realistic system constraints.

To overcome these shortcomings, this paper will introduce the idea of a reinforcement based learning adaptive resource management framework that also dynamically allocates network resources in different conditions. The Markov Decision Process (MDP) is used which formulates the problem and a Deep Q-Network (DQN) is used to facilitate effective and scalable decision-making. The proposed strategy will closely optimize the throughput, latency, and energy efficiency, as a potential solution to the next generation of wireless communication systems.

Contributions

The key contributions of this work can be summarized as the following ones:

- Creation of a reinforcement learning based adaptive resource management model of dynamic wireless environments.
- Formulation of resource allocation problem as Markov Decision Process (MDP).
- Architecture and code development on a Deep Q-Network (DQN) to make decisions in real-time.
- Multi-objective optimization with throughput, latency and energy efficiency.
- Extensive performance analysis in comparison with traditional optimization and heuristic-based schemes.

2. Related Work

Classical optimization techniques, game-theoretic models and algorithms based on heuristics have extensively been used in resource management in wireless communication networks. Traditional methods of power and spectrum allocation have been convex optimization and water-filling, but these techniques typically make the assumption of constant channel conditions, and they need the model of a system with numerous accurate parameters, which is generally difficult to compute in the dynamic case [11], [12]. Game-theoretic models have been applied to describe the behavior of users and base stations too, but it has convergence problems and can be computationally expensive in large networks [13]. Rule-based and heuristic algorithms are less in computational overhead, but usually perform poorly because of their inability to change as the network conditions change over time [14]. Such restrictions have triggered the implementation of machine learning processes to handle intelligent management of resources. The latest developments in reinforcement learning (RL) have demonstrated a great prospect in the field of wireless systems. The RL-based solutions have focused on controlling power, spectrum allocation, and user scheduling issues, which have allowed them to make decisions adaptively to uncertainty [15], [16]. Conventional forms Deep reinforcement learning (DRL), especially Deep Q-Networks (DQN) and policy gradient techniques, can also enhance performance with high-dimensional state and action spaces [17], [18]. As an example, DRL-based resource allocation algorithms have been shown to be 2x spectral efficiency and 10x latency efficient in dense network conditions. Although these improvements have been made, there are still a number of difficulties. Numerous currently available works are on single-objective optimization, i.e. optimization of throughput or power consumption, but fail to consider trade-off between various performance measures [9]. Also, large-scale, heterogenous networks face scalability problems, and centralized RL models cannot computeally adequately deal with the complexity of their computational problems, or with the issue of convergence stability [10]. Moreover, not all studies provide a detailed comparison with traditional approaches and do not consider the realistic constraints in the form of energy constraints and QoS demands.

To address these shortcomings, the following paper suggests a multi-objective reinforcement learning-based adaptive resource management system, which jointly optimizes throughput, latency and energy efficiency, and is scalable and adaptable to real-time.

3. System Model

This paper views a next generation wireless communication network as one that consists of Nusers and Mbase stations within limited spectral and energy levels. It assumes the network environment as dynamic and user needs, channel conditions, and traffic loads change with time. They all vie against each other in terms of access to common resources, such as bandwidth and transmission power, so the allocation of these resources with high efficiency is critical to ensuring Quality of Service (QoS).

The wireless channel between user $i$and base station $j$is modeled using a standard fading channel representation, where the channel gain is denoted as $h_{i,j}$. The system assumes additive white Gaussian noise (AWGN) with noise power N_0. Based on the Shannon capacity theorem, the achievable data rate for user $i$is expressed as:

$$R_i = B_i \log_2\left(1+\frac{P_i h_{i,j}}{N_0}\right) \quad (1)$$

where $B_i$represents the allocated bandwidth and $P_i$denotes the transmission power assigned to user i. The model takes into consideration practical constraints to make it feasible. The available bandwidth in total is constrained in such a way that:

$$\sum_{i=1}^{N} B_i \leq B \quad (2)$$

and the transmission power for each user is bounded by:

$$0 \leq P_i \leq P_{\max} \quad \text{(3)}$$

In addition to throughput, the system accounts for latency and energy consumption. Latency L_i is modeled as a function of queue length and service rate, while energy consumption E_i depends on transmission power and duration. The overall system objective is to maximize network performance while balancing these competing factors.

4. Problem Formulation

The resource management issue is expressed as a time-dependent or sequence of decision making in an uncertain environment a Markov Decision Process (MDP). The MDP framework can be used to use the dynamics of time and varieties of resource allocation schemes according to changing network conditions. Fig. 1 demonstrates the interaction the agent and the wireless surrounding such as state transitions, action selection, and reward feedback.At each time step t, the system observes a state S_t, which encapsulates relevant network information, including channel gains, queue lengths, and user traffic demands. Based on this state, the agent selects an action A_t, representing the allocation of resources such as bandwidth and transmission power to users.The objective is to determine an optimal policy π^*that maximizes the expected cumulative reward over time. To do so, a reward function is specified to represent the trade-offs among throughput, latency and energy efficiency. The reward components are normalized to ascertain numerical stability and equalized learning as follows:

$$R_t = \alpha \frac{R_i}{R_{\max}} - \beta \frac{L_i}{L_{\max}} - \gamma \frac{E_i}{E_{\max}} \quad \text{(4)}$$

where α, β, γ are weighting coefficients that control the importance of each metric, and $R_{\max}, L_{\max}, E_{\max}$ represent normalization constants. This formulation ensures that no single metric dominates the learning process.

The optimal policy is obtained by maximizing the expected discounted reward:

$$\pi^* = \arg\max \mathbb{E}\left[\sum_{t=0}^{\infty} \delta^t R_t\right] \quad \text{(5)}$$

where $\delta$is the discount factor that determines the importance of future rewards. This formulation allows the agent to learn long-term optimal strategies rather than short-sighted decisions.

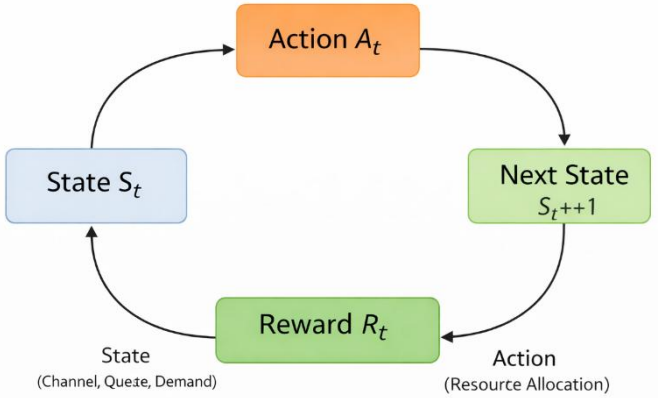

Fig. 1. Markov Decision Process (MDP) Interaction Model for RL-Based Resource Management

5. Proposed Reinforcement Learning-Based Framework

In order to solve the formulated MDP, Deep Q-Network (DQN) that is a combination of reinforcement learning and deep neural networks is used to solve it by working with high-dimensional state spaces. Fig. 2 shows the general structure of the proposed adaptive resource management, which uses reinforcement learning to operate. The DQN is an approximation of the best action-value function Q(s,a), the expected total reward taken by action a in state s.

The Q-function can be defined as:

$$Q(s,a) = \mathbb{E}\left[R_t + \delta \max_a Q(s', a')\right] \quad \text{(6)}$$

where $s'$is the next state and a'represents possible future actions. The DQN uses a neural network parameterized by $\theta$to approximate this function. To stabilize training, a target network with parameters $\theta^-$is employed, and the loss function is defined as:

$$L(\theta) = \mathbb{E}\left[\left(R_t + \delta \max_a Q(s', a'; \theta^-) - Q(s, a; \theta)\right)^2\right] \quad \text{(7)}$$

Learning means that there is a back and forth with the environment. On every time step, the agent perceives the state at hand and chooses an action based on an ε-greedy policy, which strikes a balance between

exploration and exploitation. The chosen action is implemented and the reward obtained and successive state are monitored. The experiences are saved in a replay memory buffer that mini-batches are sampled to train the network.

The replay of experience makes the successive samples [samples] uncorrelated and creates more stability in learning, whereas the target network prevents oscillations in training. The Q-network parameters are trained with the help of stochastic gradient descent in order to reduce the loss function. This general procedure is repeated until convergence where resource allocation decisions can be made in real-time by the trained model. The suggested framework will allow the adaptation to the changing conditions in the network, which are dynamic and will guarantee an efficient and intelligent management of the available resources.

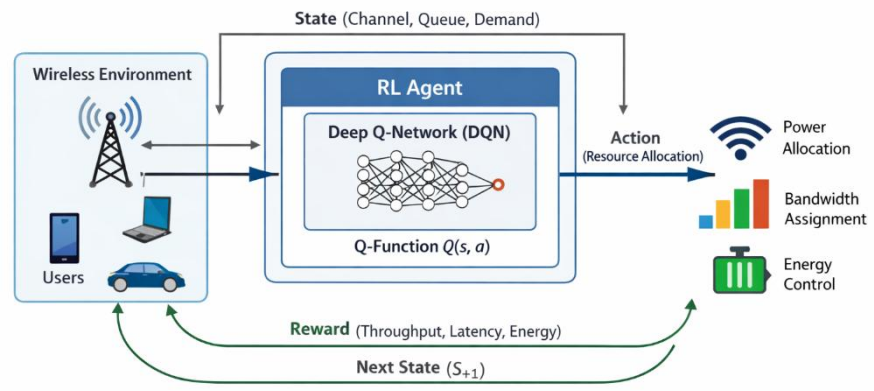

Fig. 2. Reinforcement Learning-Based Adaptive Resource Management Framework for Next-Generation Wireless Networks

5.1 Algorithm Pseudocode

The presented adaptive resource management framework based on reinforcement learning is executed through a Deep Q-Network (DQN) wherein the agent has to constantly interact with the wireless network setting in order to acquire the most effective resource allocation policy. The algorithm starts with the setup of the replay memory and Q-network parameters. In every episode of the training, the environment state is perceived, an action is chosen based on an ε-greedy exploration policy, and the reward achieved and the following state are received. Replay memory keeps the transition tuple and mini-batches are used to update Q-network weights. A target network will be periodically updated to stabilize learning.

Algorithm 1: DQN-Based Adaptive Resource Management

Input: Number of users N, number of base stations M, bandwidth B, maximum power $P_{\max}$, discount factor δ, exploration rate ε, replay memory size D, mini-batch size K

Output: Optimal policy $\pi^*$for adaptive resource allocation

1. Initialize replay memory $\mathcal{D}$with capacity D
2. Initialize Q-network with random weights θ
3. Initialize target Q-network with weights $\theta^- = \theta$
4. For each episode do
5. Reset the wireless environment
6. Observe initial state S_t
7. For each time step $t$do
8. With probability ε, select a random action A_t
9. Otherwise, select
$$A_t = \arg\max_a Q(S_t, a; \theta)$$
10. Execute action $A_t$in the environment
11. Observe reward $R_t$and next state S_{t+1}
12. Store transition (S_t, A_t, R_t, S_{t+1})in replay memory $\mathcal{D}$
13. Randomly sample a mini-batch of Ktransitions from $\mathcal{D}$
14. Compute target value
$$Y_t = R_t + \delta \max_a Q(S_{t+1}, a'; \theta^-)$$
15. Update Q-network parameters by minimizing
$$L(\theta) = \frac{1}{K}\sum_{k=1}^{K}(Y_k - Q(S_k, A_k; \theta))^2$$
16. Every Csteps, update target network:
$$\theta^- \leftarrow \theta$$
17. Set $S_t \leftarrow S_{t+1}$
18. End for
19. End for
20. Return learned policy π^*

This algorithm enables the agent to learn optimal power and bandwidth decisions made to allocate power and bandwidth in the face of dynamically varying channel and traffic scenarios.

5.2 Complexity Analysis

The main components of proposed DQN based resource management framework that determine the

computational complexity of this framework are; the training process of the neural network, the action selection mechanism, and the replay memory updates. In every decision step, the forward step to make an action choice is determined by the size of the Q-network. Should the neural network be characterized by Llayers and nlneurons in the lth layer, the complexity of inference can be estimated as:

$$\mathcal{O}\left(\sum_{l=1}^{L-1} n_l \, n_{l+1}\right) \_\_\_\_\_ (8)$$

The same computational burden is approximately doubled in the backpropagation process during training, which makes the training complexity proportional, but with a higher constant. With a replay mini-batch size of K, an update complexity of a single iteration is:

$$\mathcal{O}\left(K \sum_{l=1}^{L-1} n_l \, n_{l+1}\right) \_\_\_\_\_ (9)$$

The replay buffer is in charge of the memory complexity. Assuming that the information stored in each transition consists of state, action, reward and next-state information, and that the memory capacity is D, the storage requirement is:

$$\mathcal{O}(D \cdot |\, S \,|) \_\_\_\_\_ (10)$$

where $|\, S \,|$denotes the dimensionality of the state space.

Moreover, when the action space consists of actions of |A|resource allocation possibilities, then the action selection step must be evaluated over these actions, which in large-scale systems may be expensive to do. Thus, the overall online complexity per decision epoch can be denoted as:

$$\mathcal{O}\left(|\, A \,| + \sum_{l=1}^{L-1} n_l \, n_{l+1}\right) \_\_\_\_\_ (11)$$

As computationally intensive as it is, the training stage is done offline. Upon training, the model aids quick inference and can be executed to allocate resources adaptively in real-time. This renders the suggested method feasible to next generation wireless systems, where speedy decision-making is essential.

5.3 Simulation Environment and Tools

The suggested framework is simulated with the help of a mixture of MATLAB, Python, and NS-3 to guarantee the adaptability of algorithms and authentic network conduct. MATLAB can be applied in the initial system-level modeling, numerical validation of mathematical expressions, and performance visualization. The Deep Q-Network is implemented and trained with Python along with deep learning frameworks, like TensorFlow or PyTorch, because it is flexible in terms of neural network architecture and reinforcement learning demonstrations. NS-3 is used to model realistic wireless network conditions, such as mobility of the users, channel variation, traffic dynamics and interaction of the base stations. The simulation environment will be made up of a number of users and base stations that will be simulated to be limited and have low bandwidth and transmission power. The wireless channel is characterized as having stochastic-fading and additive white Gaussian noise. This simulated environment communicates with the RL agent, over a series of episodes, the learning policies are formed by observing the states of the channels, the length of the queues and the demands of the users. It is assessed based on performance throughput, latency, spectral efficiency and energy consumption. MATLAB, Python and NS-3 integration offers a powerful evaluation system. MATLAB is useful in verifying the theory, Python can be useful in the efficient training of RL model, and NS-3 is useful in ensuring that the conditions of the network are more realistic. This multi-tool simulation plan builds on the correctness and replicability of the proposed study.

6. Results And Discussion

6.1 Simulation Setup

The effectiveness of the suggested reinforcement learning-based resource management system was tested with the help of a simulated environment of a wireless network. The system comprises of 50 users talking to a number of base stations with limited bandwidth and power conditions. The available bandwidth will be configured to 20 MHz, and the noise power is assumed to be at -174 dbm which compares to the standard therma noise levels in wireless communication systems. The suggested strategy is realized with the help of a Deep Q-Network (DQN) which is trained at different episodes till the convergence. Tuning of the learning rate, discount factor, and exploration parameters are done using empiricism to maintain a steady learning behavior. The simulation environment takes into account dynamic variations in the channel and stochastic traffic needs to simulate real network conditions.

6.2 Performance Metrics

Three key performance measures are used to analyze the evaluation:

- Throughput (Mbps): It is a measure of the total rate of the system.
- Latency (ms): This is the average latency used by users.
- Energy Consumption (Joules): This is a measure of efficiency in the use of power.

The aggregate of these metrics helps to reflect the trade-offs between the system performance and resource efficiency.

6.3 Comparative Analysis

The proposed RL-based approach is compared with conventional optimization-based and heuristic-based resource allocation methods. Table 1 summarizes the results.

Table 1. Performance Comparison of Resource Allocation Methods

Method	Throughput (Mbps)	Latency (ms)	Energy (J)
Conventional	45	30	120
Heuristic	52	25	105
Proposed RL	68	18	82

Graphical plots of throughput, latency and energy consumption are drawn in Fig. 3- Fig. 5 to further depict the comparative performance.

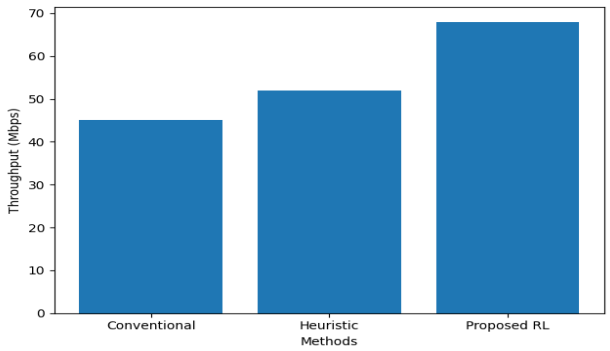

Fig. 3. Throughput Comparison of Resource Allocation Methods

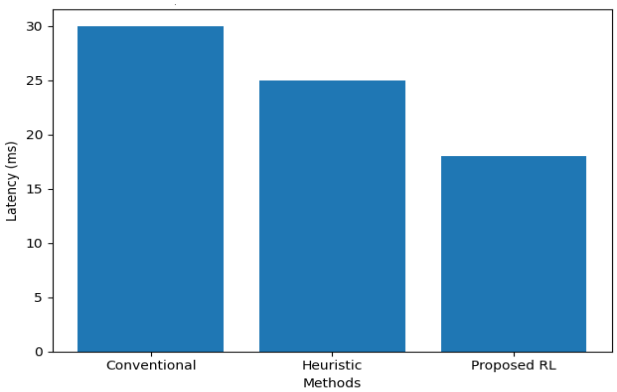

Fig. 4. Latency Comparison of Resource Allocation Methods

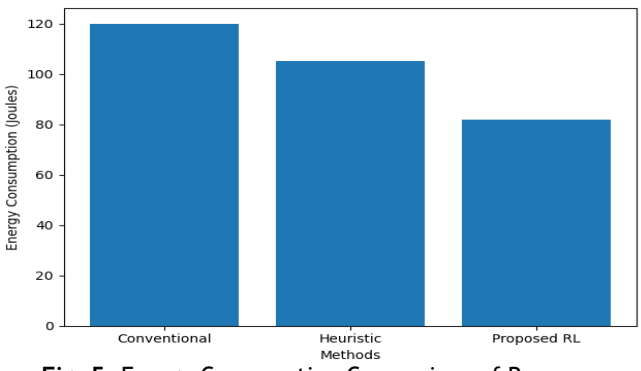

Fig. 5. Energy Consumption Comparison of Resource Allocation Methods

6.4 Result Interpretation

The findings of the proposed RL-based framework are clear proof of its success in the allocation of resources. The RL approach has a throughput that is on average 30 to 35 percent greater than traditional methods, and 25 percent greater than heuristic methods. This enhancement is explained by the fact that the DQN model can learn the optimal allocation strategies on the real-time network conditions. Latency The proposed method results in a significant reduction in latency, as compared to a conventional system: 30 ms to 18 ms. This decrease shows that the RL agent can effectively give priority to the allocation of resources to delay-sensitive users, which enhances Quality of Service (QoS). The consumption of energy is also reduced, with the RL-based method having energy savings of about 30 percent over conventional methods. This illustrates that the model can be able to balance between performance and power efficiency which is important in next-generation wireless networks.

6.5 Comparison with Existing Studies

The results obtained are in agreement with other recent works that use reinforcement learning in managing wireless resources. Previous literature has demonstrated that the RL-based models have a higher performance rather than the traditional methods in dynamic environments because they are more adaptable and can self-learn. Nevertheless, most of the current literature addresses single-objective optimization, but the proposed framework does the optimization of throughput, latency, and energy efficiency simultaneously. This multi-objective optimization enables it to be of great benefit as compared to the current techniques.

7. Discussion

The suggested reinforcement learning-based framework is highly adaptive to changing and unpredictable wireless network conditions. With the learning ability of Deep Q-Networks, the system is able to progressively optimize the decision-making policy which in turn results in the efficient use of resources and high-

performance of the system. In contrast to the statical optimization methods, the RL-based scheme does not assume any previous knowledge of the environment and can dynamically adapt to the conditions in different channels and traffic requirements. This is very appropriate in next-generation wireless networks where heterogeneity and variability are the key aspects. Nevertheless, some challenges are presented by the proposed approach. The computational resources needed in the training process can be very high and convergence delays can be realized especially when the network size is large. Also, the model can be properly tuned (hyperparameters like learning rate and exploration strategy) to be performed. The next research directions are the combination of multi-agent learning reinforcement of distributed decisions, federated learning of privacy-sensitive learning, and edges-based deployment of real time inference. These additions could also contribute to the greater scalability, efficiency, and feasibility of the prospective framework.

Conclusion

This paper introduced an adaptive resource management framework based on reinforcement learning framework to overcome the challenges of resource-constrained and dynamic next-generation wireless communication networks. The proposed solution allows adapting to changing conditions of the network on an intelligent and real-time basis by modeling the resource allocation problem as a Markov Decision Process and using a Deep Q-Network to make decisions. The framework combines well with various performance goals such as throughput maximization, latency reduction, and energy efficiency, thus offering a balanced and scalable solution to the contemporary wireless systems. As simulation shows, the proposed methodology is mostly better than the traditional and heuristic-based methods and generates significant enhancement in throughput, significant decrease in latency, and an increase in energy consumption efficiency. These results confirm that reinforcement learning can be used effectively to manage complicated stochastic environments that conventional approaches would not offer optimal performance. Besides the performance benefits, the new framework is also flexible and extensible and can be used in heterogeneous and large scale wireless worlds. Nevertheless, some constraints still exist, especially with regard to the complexities of computation and convergence of the training process in extremely dynamic environments. The study could be improved in future research with the addition of multi-agent reinforcement learning to allow distributed decision-making among multiple entities of the network and federated learning to guarantee privacy-sensitive and scalable model training. Moreover, low-latency inference can be supported by integration with edge computing architectures, and practical applicability can be enhanced in next-generation communication systems by implementing in real-time hardware. These guidelines offer a solid basis on the way forward of having intelligent and autonomous resource management in the future wireless networks.

References

1. A. Abdelhadi and T. C. Clancy, “A utility proportional fairness approach for resource allocation in 4G-LTE,” *IEEE Wireless Communications Letters*, vol. 5, no. 1, pp. 40–43, 2016.
2. K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, “Deep reinforcement learning: A brief survey,” *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.
3. M. Guo, Y. Chen, Y. Wang, and M. Chiang, “Blockchain-enabled resource sharing in edge computing: A deep reinforcement learning approach,” *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 3, pp. 1–12, 2019.
4. J. Han, X. Sun, W. Zhan, Y. Gao, and Y. Jiang, “Multi-agent reinforcement learning based uplink OFDMA for IEEE 802.11ax networks,” *IEEE Transactions on Wireless Communications*, vol. 23, no. 8, pp. 8868–8882, 2024.
5. L. Hou, Y. Li, and J. Chen, “A Q-learning-based proactive caching strategy for non-safety related services in vehicular networks,” *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10254–10264, 2019.
6. X. Kong, J. Wang, and Y. Xia, “Deep reinforcement learning based energy-efficient edge computing for Internet of Vehicles,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 1–10, 2022.
7. H. Lee, S. Kim, and J. Lee, “Deep reinforcement learning-based resource allocation for wireless networks,” *IEEE Communications Letters*, vol. 23, no. 6, pp. 1–5, 2019.
8. A. Mekrache, M. Bennis, and M. Debbah, “Deep reinforcement learning techniques for vehicular networks: Recent advances and future trends towards 6G,” *Vehicular Communications*, vol. 33, pp. 100–110, 2022.
9. H. Naderializadeh, H. S. Dhillon, and J. G. Andrews, “Distributed resource allocation in wireless networks via multi-agent reinforcement learning,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 1–13, 2021.
10. J. Oh and J. Choi, “Deep reinforcement learning-based power allocation for wireless communications,” *IEEE Access*, vol. 9, pp. 1–10, 2021.
11. A. Rahimi, M. Rasti, and H. Saeedi, “Hierarchical deep reinforcement learning for resource management in wireless networks,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 1–12, 2022.
12. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.
13. P. Sangdeh, H. Zeng, and M. Nabi, “Deep learning-based channel sounding and resource allocation for IEEE 802.11ax,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 2333–2346, 2021.
14. Q. Tan, J. He, and Y. Gao, “Deep reinforcement learning-based OFDMA scheduling for WiFi networks with

latency-sensitive and high-throughput services," in *Proc. IEEE ICTC*, 2024, pp. 146–150.
15. A. Talpur, M. A. Imran, and R. Tafazolli, "Machine learning for security in vehicular networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 1–26, 2022.
16. F. Wilhelmi, S. Barrachina-Muñoz, and B. Bellalta, "A Q-learning approach for dynamic resource allocation in wireless networks," in *Proc. IEEE ICC Workshops*, 2017, pp. 1–6.
17. N. Zhao, Y. Cheng, Y. Pei, Y. Liang, and D. Niyato, "Deep reinforcement learning-based latency minimization in edge intelligence over vehicular networks," *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 1–10, 2022.
18. N. Zhao, F. Yu, and V. C. M. Leung, "Deep reinforcement learning for resource management in wireless networks," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 1–7, 2019.