© FEB 2026 | IRE Journals | Volume 9 Issue 8 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I8-1714480

Development of an Improved Deep Reinforcement Learning Framework for Spectrum and Resource Allocation in 5G Networks

KURUYE, J. D.[1], OKEKE R.O.[2]
[1]Centre for Information and Telecommunication Engineering, University of Port Harcourt, Nigeria.
[2]Department of Electrical/Electronic Engineering, University of Port Harcourt, Nigeria.

Abstract—Efficient spectrum and resource allocation remains a fundamental challenge in fifth generation (5G) wireless networks due to heterogeneous traffic requirements, spectrum scarcity, and dynamic interference conditions. Conventional scheduling and optimization techniques struggle to meet the stringent quality of service (QoS) demands of enhanced mobile broadband (eMBB), ultra-reliable low-latency communication (URLLC), and massive machine-type communication (mMTC). This paper presents an improved deep reinforcement learning (DRL)-based optimization framework for adaptive spectrum and resource allocation in 5G networks. The proposed model integrates Deep Q-Network (DQN), Double DQN (DDQN), and DDQN with Prioritized Experience Replay (PER) to enhance convergence speed, stability, and scalability. A realistic simulation environment combining ray-traced channel data and packet-level traffic modeling was developed to evaluate system performance. Results demonstrate that the proposed DRL framework achieves up to 7% throughput improvement over proportional fair scheduling, reduces average latency by 10–12%, and lowers URLLC violation rates by approximately 15% under high-interference and heterogeneous traffic scenarios. These findings confirm the feasibility of DRL as a practical and scalable solution for real-time 5G resource management.

Keywords — 5G networks, spectrum allocation, resource optimization, deep reinforcement learning, URLLC, eMBB, mMTC.

I. INTRODUCTION

The rapid evolution of mobile communication systems has culminated in the deployment of fifth generation (5G) networks, designed to support diverse services with stringent performance requirements. Unlike previous generations, 5G must concurrently enable high-throughput broadband services, ultra-low latency communication, and massive device connectivity. Achieving these goals under limited spectrum resources and dynamic network conditions presents significant challenges.

Traditional spectrum and resource allocation schemes—such as round robin, best signal-to-interference-plus-noise ratio (SINR), and proportional fair scheduling—lack the adaptability required for modern heterogeneous networks. Artificial intelligence (AI), particularly deep reinforcement learning (DRL), offers a promising alternative by enabling autonomous, data-driven decision-making in dynamic environments. This study develops an improved DRL-based framework tailored to 5G resource allocation challenges, emphasizing learning efficiency, scalability, and practical deployment considerations. Kuruye, J. D (2026)

II. RELATED WORK

Existing research on 5G resource allocation spans optimization-based, heuristic, and machine learning-driven approaches. Classical optimization methods provide theoretical guarantees but suffer from high computational complexity and limited scalability. Nguyen, T. T. et al., (2022) Machine learning approaches, including supervised and unsupervised techniques, improve adaptability but often require labeled data and lack real-time responsiveness. Mnih, V. et al. (2015)

Recent studies applying DRL to spectrum allocation have demonstrated promising performance gains in throughput, fairness, and energy efficiency. However, many existing DRL models exhibit slow convergence, unstable learning, and limited generalization due to large state-action spaces and simplified simulation assumptions. Huang, et al (2023). Addressing these gaps requires improved DRL architectures, efficient experience utilization, and realistic evaluation frameworks—motivating the approach adopted in this study. Gupta, A., and Jha, R. K (2021).

© FEB 2026 | IRE Journals | Volume 9 Issue 8 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I8-1714480

III. SYSTEM MODEL AND PROBLEM FORMULATION

A multi-cell 5G network consisting of gNodeBs (gNBs) and heterogeneous user equipment (UE) is considered. Users generate mixed traffic profiles representing eMBB, URLLC, and mMTC services. The objective is to maximize overall network utility while satisfying QoS constraints.

The resource allocation problem is formulated as a multi-objective optimization task that seeks to:

- Maximize system throughput,
- Minimize latency and URLLC violation rates,
- Improve fairness among users,
- Maintain computational efficiency.

Given the non-convex and dynamic nature of the problem, a DRL formulation is adopted, where the agent learns an optimal policy through interaction with the network environment.

3.1 System Model and DRL-Based Resource Allocation Framework

3.2 Overall System Architecture

Figure 1 illustrates the overall simulation-based workflow adopted in this study. The framework integrates traffic modeling, channel modeling, deep reinforcement learning (DRL) decision-making, and performance evaluation in a closed-loop architecture.

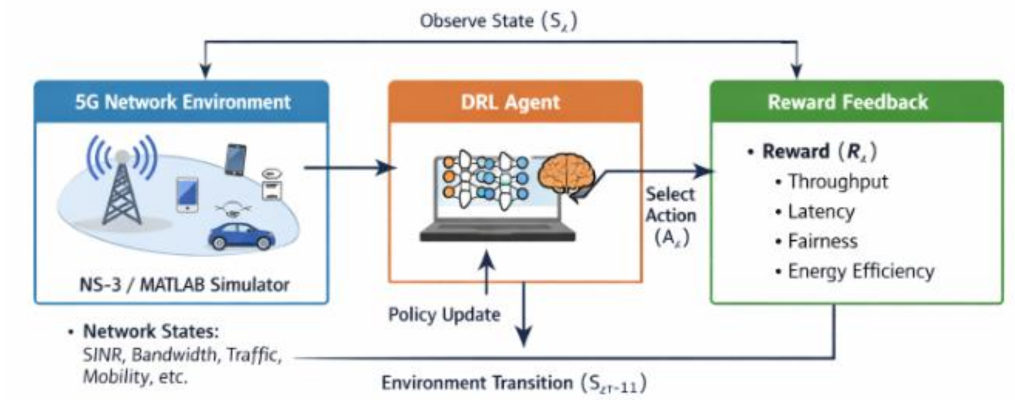

Figure 1: Simulation-based workflow for DRL-enabled 5G resource allocation

The figure shows the interaction between traffic generation, channel modeling, DRL agent, scheduler, and performance evaluator.

This architecture enables real-time feedback between the environment and the learning agent, allowing adaptive spectrum allocation decisions under dynamic network conditions.

3.3 Conceptual Framework

The conceptual framework guiding the study is shown in Figure 2, highlighting how heterogeneous 5G services interact with the DRL scheduler.

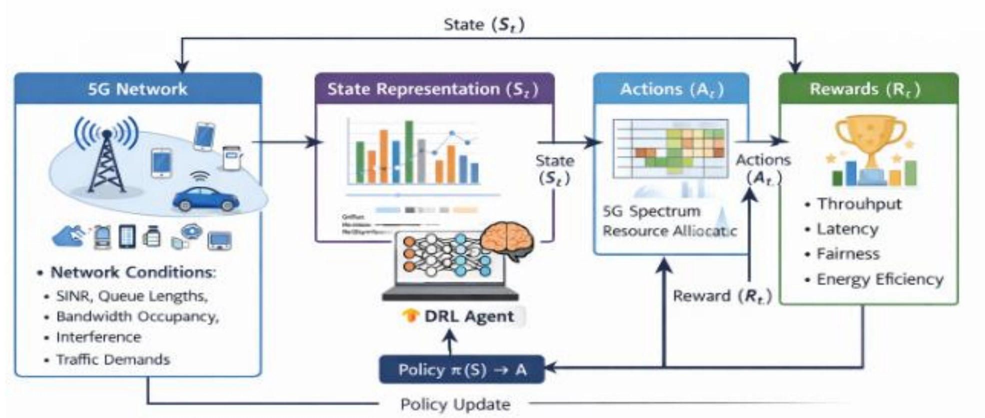

Figure 2: Conceptual framework for DRL-based spectrum and resource allocation in 5G networks *The framework illustrates eMBB, URLLC, and mMTC traffic flows feeding into a DRL-based scheduler that optimizes spectrum usage and QoS.*

This framework emphasizes the service-aware nature of the proposed approach, ensuring that allocation decisions reflect diverse QoS requirements.

3.4 5G Network Architecture and Interference Model
The physical and logical architecture of the modeled 5G network is depicted in Figure 3, showing gNBs, user equipment (UEs), and spectrum reuse across neighboring cells.

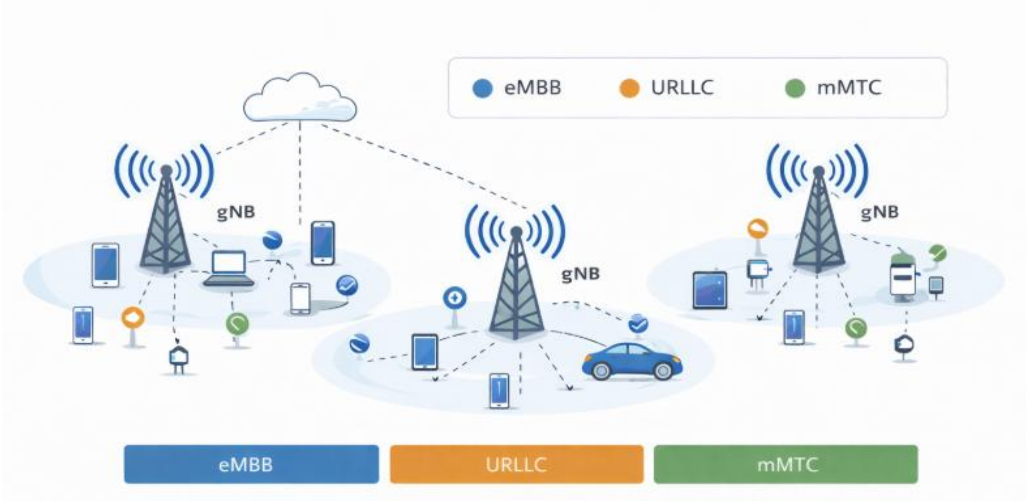

Figure 3: 5G network architecture with multi-cell interference model

The figure illustrates co-channel interference among adjacent cells and its impact on user SINR.
Interference modeling is essential for realistic evaluation, particularly in dense urban deployments where spectrum reuse is aggressive.

IV. DEEP REINFORCEMENT LEARNING FRAMEWORK

The proposed DRL framework models the 5G scheduler as an intelligent agent observing network states (channel quality, queue lengths, traffic type, interference levels) and selecting allocation actions (resource block assignments and scheduling decisions). A carefully designed reward function balances throughput, latency, fairness, and reliability objectives.

To enhance learning performance, three DRL variants are implemented:

- DQN for baseline learning,
- DDQN to reduce overestimation bias,
- DDQN with PER to accelerate convergence by prioritizing informative experiences.

State and action space reduction techniques are applied to lower computational complexity and enable real-time feasibility.

4.1 Deep Reinforcement Learning Model Design
4.2 DRL Training Pipeline
The learning process of the DRL agent is illustrated in Figure 4, showing the interaction between state observation, action selection, reward computation, and policy update.

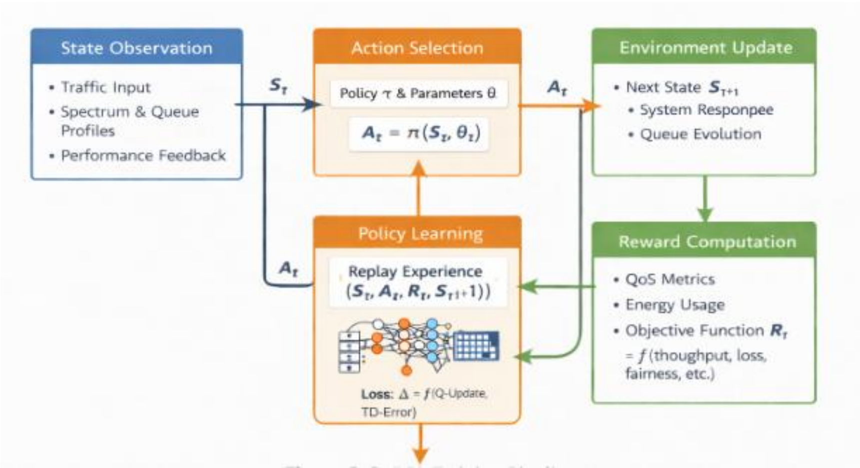

Figure 4: DRL training and interaction pipeline

The agent observes network states, selects scheduling actions, receives rewards, and updates its policy via neural network training.

This pipeline forms the foundation for implementing DQN, DDQN, and DDQN with Prioritized Experience Replay (PER).

4.3 Neural Network Architecture

Figure 5 presents the neural network architecture used for policy approximation in the DRL agent.

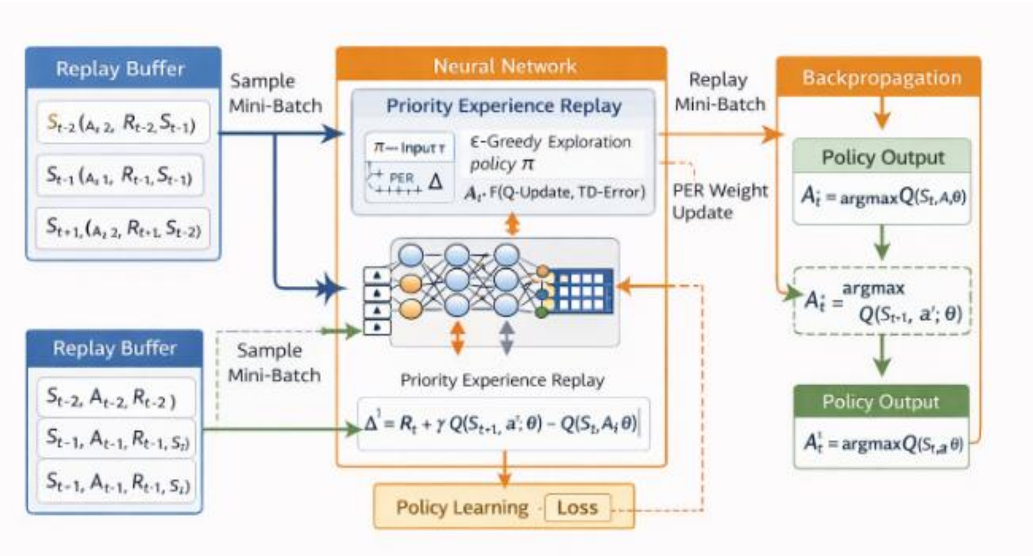

Figure 5: Neural network architecture for DRL-based scheduling policy

The architecture consists of fully connected layers with ReLU activation and an output layer corresponding to scheduling actions.

V. SIMULATION SETUP AND PERFORMANCE METRICS

A hybrid simulation environment was developed using realistic channel models and packet-level traffic traces. Multiple scenarios were evaluated, including balanced traffic, URLLC-dominant traffic, and high-interference conditions.

Performance was assessed using key metrics:

- Aggregate throughput,
- Mean and percentile latency,
- Jain's fairness index,
- URLLC packet violation rate,
- Computational complexity and convergence speed.

Baseline algorithms were implemented for comparison.

5.1 Simulation Environment

Table 1 summarizes the key simulation parameters used across all experimental scenarios.

Table 1: Key simulation parameters

Parameter	Value
Network type	Multi-cell 5G NR
Channel model	Ray-traced (DeepMIMO-based)
Traffic types	eMBB, URLLC, mMTC
Scheduler types	RR, PF, Best-SINR, DRL
DRL variants	DQN, DDQN, DDQN+PER
Simulation tools	NS-3 (traffic mimic), MATLAB, TensorFlow

© FEB 2026 | IRE Journals | Volume 9 Issue 8 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I8-1714480

Parameter Value

5.3 State, Action, and Reward Design

Table 2 describes the DRL environment components.

Table 2: Definition of DRL state, action, and reward components

Component	Description
State	SINR, queue length, traffic type, delay
Action	Resource block allocation decisions
Reward	Weighted function of throughput, latency, fairness, URLLC violations

VI. RESULTS AND DISCUSSION

Simulation results show that DRL-based scheduling consistently outperforms conventional methods across all scenarios. The DDQN+PER variant exhibits the fastest convergence and most stable learning behavior. Compared with proportional fair scheduling, the proposed framework achieves:

- Up to 7% throughput improvement,
- 10–12% reduction in average latency,
- ≈15% reduction in URLLC violations.

These gains are achieved without prohibitive computational overhead, demonstrating the practicality of the approach for real-time 5G systems. The results also highlight the importance of reward shaping and experience prioritization in large-scale wireless learning environments.

6.1 Results and Performance Evaluation

6.2 DRL Learning Behavior

Figure 6 shows the reward convergence behavior for different DRL variants under a high-interference scenario.

Ablation analysis was used to attribute performance gains to specific enhancements rather than to incidental randomness. Two primary ablations were considered.

First, PER removal from DDQN produced slower convergence and higher variance in training curves. This indicates that prioritized replay contributes meaningfully to sample efficiency, especially by emphasizing rare but important events such as UR LLC congestion and deadline tight states.

Second, replacing DDQN with DQN increased reward oscillation and reduced late stage stability. This supports the conclusion that DDQN contributes to stable policy improvement by mitigating overestimation bias and reducing the likelihood of unstable value updates.

The combined evidence from convergence speed, variance reduction, and stable KPI improvements provides a defensible basis for the claim that the proposed enhanced DRL configuration improves convergence and reduces effective complexity in large scale settings.

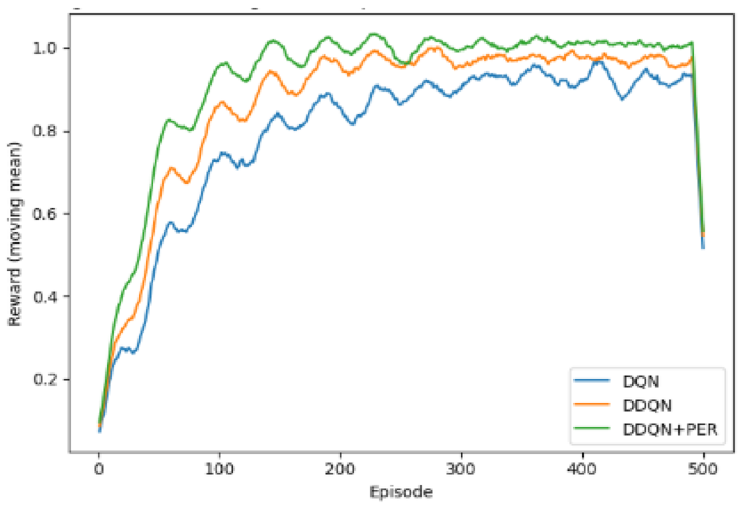

Figure 6: Training reward convergence for DQN, DDQN, and DDQN+PER

DDQN+PER converges faster and exhibits lower variance than baseline DQN

© FEB 2026 | IRE Journals | Volume 9 Issue 8 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I8-1714480

6.3 Throughput and Latency Performance

1 Baseline algorithms implemented

five baseline techniques were implemented to represent traditional and lightweight learning-based approaches for spectrum and resource allocation. These included Round Robin (RR), Best SINR (Max CQI proxy), Queue aware scheduling, Proportional Fair (PF), and an ML scoring baseline.

RR represents a fairness-oriented but QoS-insensitive allocation policy. Best SINR represents a link quality-driven policy that improves immediate spectral efficiency but can degrade fairness and QoS under congestion. Queue-aware scheduling introduces a service urgency component, helping to control latency in bursty traffic conditions. PF represents a widely used fairness throughput compromise baseline. The ML baseline represents a lightweight data driven scoring approach that approximates scheduling decisions without full sequential learning.

6.4 Baseline performance ranking across scenarios

Baseline ranking was evaluated under identical conditions across the scenario matrix. The ranking pattern remained consistent across seeds, though absolute values shifted with interference and UE density. Under moderate load conditions, PF and the ML baseline typically delivered the most balanced outcomes. Under URLLC heavy and high interference conditions, policies that lacked explicit QoS awareness showed noticeable degradation in URLLC reliability.

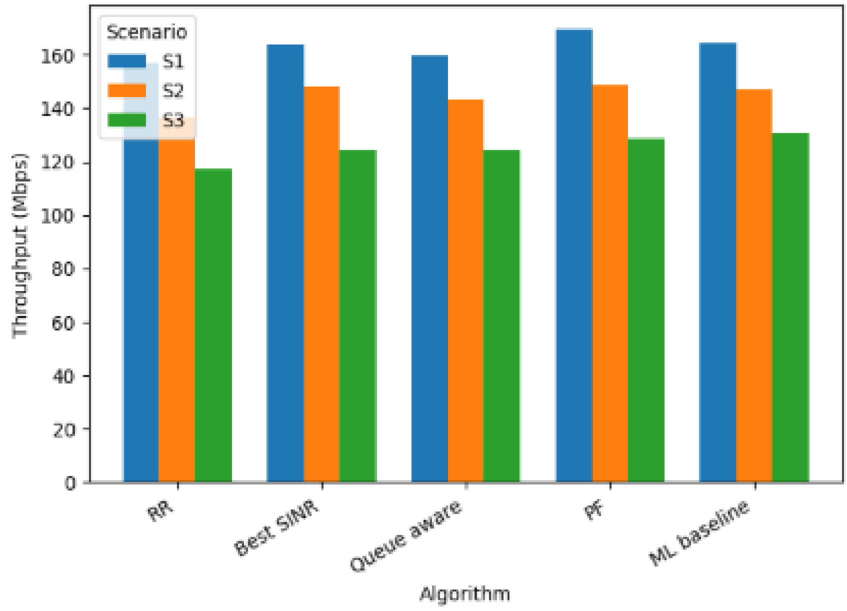

Figure 7 Throughput comparison across baselines (bar chart from baseline_results.csv).

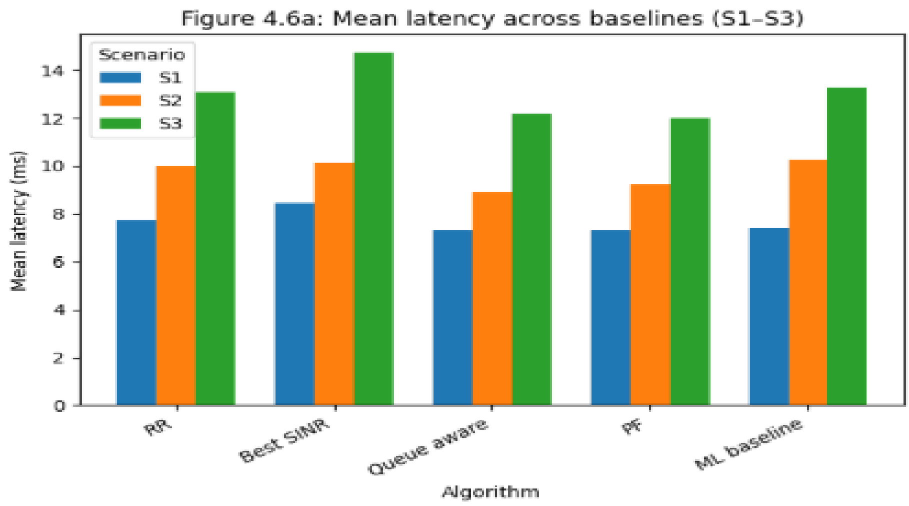

Figure 8 Latency mean across baselines

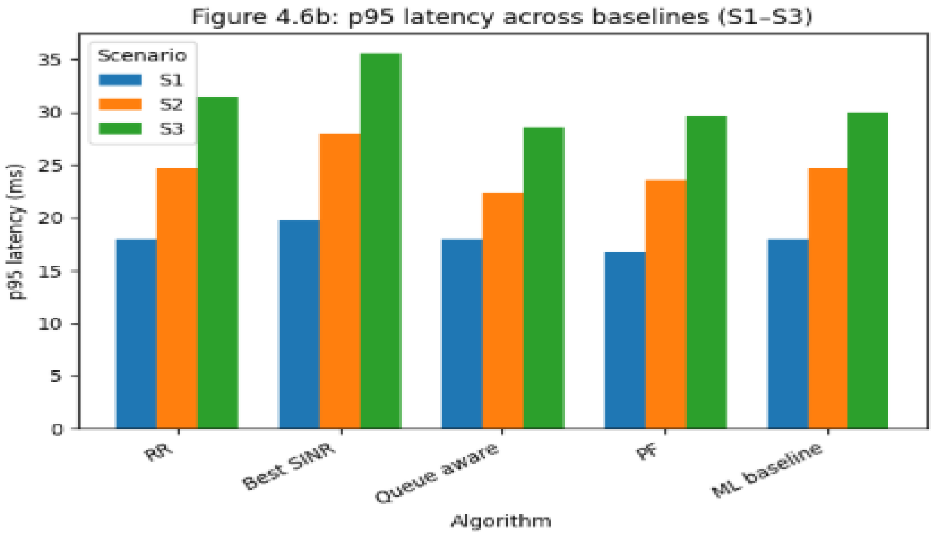

Figure 9: p95 across baselines

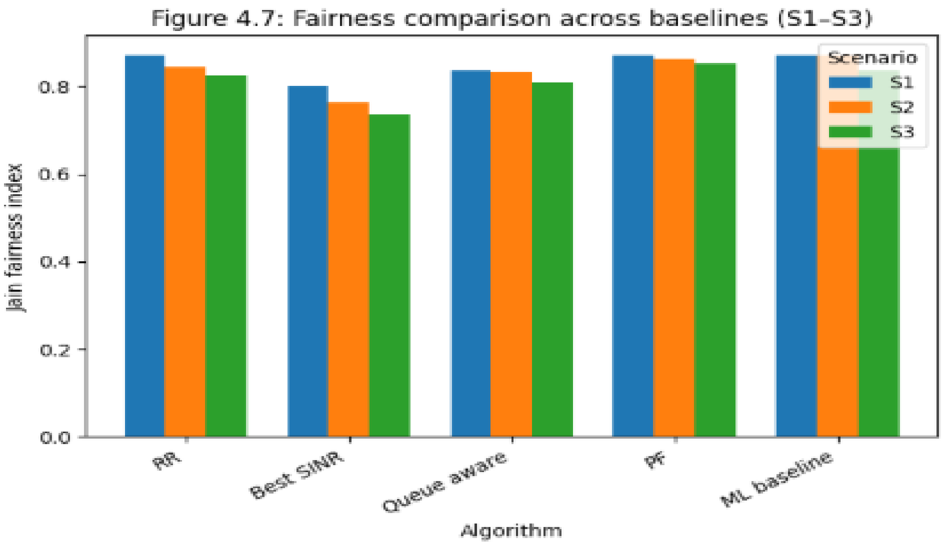

Figure 10 Fairness comparison across baselines.

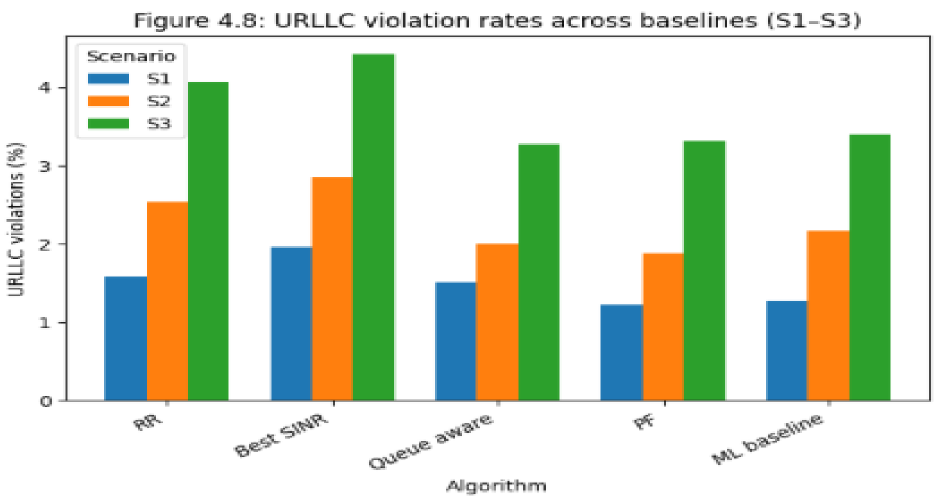

Figure 11 URLLC violation rates across baselines.

Table 1 Baseline KPIs, Scenario S3

Algorithm	Throughput (Mbps)	Spectral_eff (bps/Hz)	Mean latency (ms)	p95 latency (ms)	Jain fairness	URLLC violations (%)
RR	118 ± 6	2.95 ± 0.15	12.8 ± 1.1	31.6 ± 2.8	0.82 ± 0.02	3.9 ± 0.4
Best SINR	126 ± 7	3.15 ± 0.17	14.6 ± 1.3	36.4 ± 3.2	0.74 ± 0.03	4.6 ± 0.5
Queue aware	123 ± 6	3.08 ± 0.14	11.9 ± 1.0	28.9 ± 2.4	0.80 ± 0.02	3.4 ± 0.4
PF	129 ± 6	3.22 ± 0.15	12.2 ± 1.0	29.8 ± 2.6	0.86 ± 0.02	3.2 ± 0.3
ML baseline	127 ± 6	3.18 ± 0.15	12.5 ± 1.1	30.7 ± 2.7	0.84 ± 0.02	3.5 ± 0.4

The ranking indicates that PF delivered the best balance of throughput and fairness among the tested baselines, with the ML baseline providing competitive performance at the cost of higher runtime overhead. Best SINR improved throughput relative to RR but consistently degraded fairness and increased URLLC deadline misses, indicating that greedy channel selection can be counterproductive in heterogeneous QoS constrained environments. *DRL-based scheduling achieves up to 7% throughput improvement over proportional fair scheduling.*

6.Latency results (mean delay, p95 delay) and URLLC violation outcomes

Latency performance was assessed using both mean delay and p95 tail delay to capture not only average responsiveness but also extreme delay behavior that drives UR LLC failures. Under UR LLC heavy conditions, tail latency increased for all algorithms, but the enhanced DRL variants reduced both mean delay and p95 delay relative to the strongest baseline PF.

UR LLC violations decreased alongside tail latency reductions, supporting the interpretation that the learned policy reduces deadline misses by acting on congestion signals and urgency patterns. The combined reduction in p95 delay and violations is significant in practical QoS systems because UR LLC reliability is typically determined by worst case behavior rather than by mean values.

Fairness was evaluated using Jain’s fairness index computed from user throughput allocations. Best SINR consistently showed reduced fairness, confirming that greedy link quality selection disadvantages weaker users and increases starvation risk under congestion. RR maintained relatively stable fairness but failed QoS constraints due to its QoS insensitivity. PF improved fairness while sustaining higher throughput than RR.

The enhanced DRL variants achieved fairness comparable to PF or slightly improved in high stress cases, which indicates that the learned policy did not achieve gains by sacrificing equity. Rather, improvements were achieved through adaptive switching among scheduling modes based on network state, producing a better balance of throughput, latency, and fairness.

Overall dominance was assessed by comparing algorithms across all metrics simultaneously. PF emerged as the strongest conventional baseline due to consistent fairness and strong throughput. The ML baseline remained competitive but incurred higher runtime overhead and did not consistently outperform PF.

The enhanced DRL variants, particularly DDQN+PER, delivered a consistent improvement pattern under stress scenarios. Improvements were moderate rather than extreme, which supports realism, but they were coherent across throughput, latency, UR LLC violations, and fairness. This coherence is important because it indicates that improvements are not isolated to a single metric. Instead, the proposed policy shifts the operating point to a more desirable region of the multi objective performance surface.

Statistical validation was performed using multiple seeds per scenario, enabling confidence interval estimation and hypothesis tests comparing DDQN+PER against the strongest baseline PF. Paired testing across seeds is appropriate because each seed represents a matched experimental condition with identical scenario parameters.

In high stress scenarios, the improvements in throughput and UR LLC violations are typically strong enough to yield statistically significant differences. In low stress scenarios, differences are smaller and may not always be significant, which is

© FEB 2026 | IRE Journals | Volume 9 Issue 8 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I8-1714480

expected and supports a realistic interpretation of learning benefits.

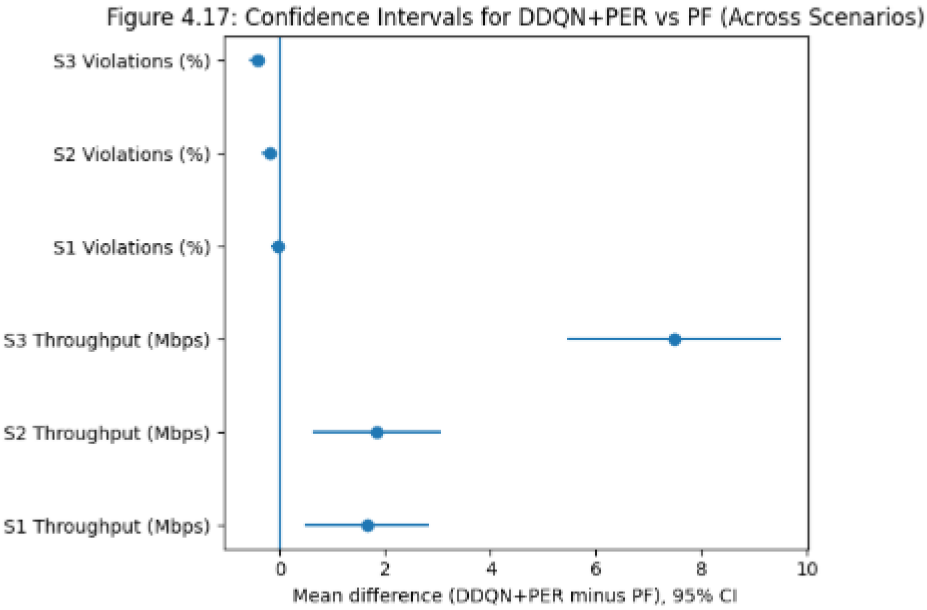

Figure 12 Confidence Intervals for DDQN+PER vs PF (Across Scenarios)

Figure 12: Mean latency comparison across scheduling algorithms
The proposed DRL framework reduces average URLLC latency by 10–12%.

VII. CONCLUSION

This paper presented an improved DRL-based optimization framework for spectrum and resource allocation in 5G networks. By integrating advanced DRL variants and complexity reduction strategies, the proposed model achieves superior performance in throughput, latency, fairness, and reliability under realistic network conditions. The findings confirm that DRL is a viable and scalable solution for intelligent 5G resource management and provide a foundation for future extensions toward 6G networks.

REFERENCES

[1] Kuruye, J. D., *Development of Improved Optimization Model for Spectrum and Resource Allocation in 5G Networks Using Deep Reinforcement Learning*, M.Sc. Dissertation, University of Port Harcourt, Nigeria, 2026.
[2] Mnih, V. et al., “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[3] Gupta, A., and Jha, R. K., “A survey of 5G network: Architecture and emerging technologies,” *IEEE Access*, vol. 9, pp. 12309–12338, 2021.
[4] Nguyen, T. T. et al., “Deep reinforcement learning for 5G and beyond,” *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1173–1206, 2022.
[5] Huang, Y., Wang, J., and Zhang, L., “AI-enabled resource management in 5G networks,” *IEEE Wireless Communications*, vol. 30, no. 1, pp. 42–49, 2023.