BENCHMARKING CLASSICAL VS QUANTUM REINFORCEMENT LEARNING

A Comparative Analysis of Classical and Quantum-Enhanced RL Models

HAROON RACHEL EVA

NOVEMBER 2025

Abstract

This report presents a comparative study between classical reinforcement learning (RL) algorithms (DQN, PPO, DDPG, TD3) and quantum-enhanced RL models (Quantum Actor-Critic, Quantum Policy Gradient, Quantum Q-Network). The experiments were conducted across discrete and continuous control environments using OpenAI Gym. Classical RL algorithms demonstrated strong convergence and policy stability, while quantum-hybrid models exhibited promising early-stage exploration and faster reward improvement. The goal of this report is to establish a foundational benchmark highlighting how quantum computation can enhance reinforcement learning performance in the near future.

Contents

1. INTRODUCTION

1.1. Overview

Reinforcement Learning (RL) enables agents to learn through interaction with their environment. Over the last decade, RL has achieved success in diverse fields such as robotics, gaming, finance, and autonomous systems. However, classical RL faces challenges with scalability, sample inefficiency, and convergence instability in high-dimensional or continuous environments.

Quantum Reinforcement Learning (QRL) leverages quantum computation principles like superposition and entanglement to improve state-space representation and exploration efficiency. Quantum-enhanced RL models combine quantum circuits with classical learning frameworks to explore faster convergence and better policy diversity.

1.2. Objective

This study aims to:

- Benchmark major classical RL algorithms on standard Gym environments.
- Evaluate quantum-enhanced RL models on similar tasks.
- Compare convergence patterns, stability, and performance metrics.

1.3. Report Structure

The report is divided into the following sections:

- Section 2: Classical Reinforcement Learning.
- Section 3: Quantum Reinforcement Learning.
- Section 4: Comparative Discussion.
- Section 5: Conclusion and Future Scope.

2. CLASSICAL REINFORCEMENT LEARNING

2.1. Overview

Classical RL algorithms were trained using the `Stable-Baselines3` library on four Gym environments—`CartPole-v1`, `LunarLander-v2`, `Pendulum-v1`, and `BipedalWalker-v3`. Each model ran for 100 episodes, and key metrics such as mean reward, median reward, and episode length were recorded.

2.2. Algorithms

- **Deep Q-Network (DQN):** Uses value approximation to estimate action-value pairs for discrete environments.
- **Proximal Policy Optimization (PPO):** Policy-gradient-based method ensuring stable updates.
- **Deep Deterministic Policy Gradient (DDPG):** Continuous control using an actor-critic setup.
- **Twin Delayed DDPG (TD3):** An improved version of DDPG reducing overestimation bias.

2.3. Performance Summary

Table 1: Aggregated Classical RL Performance Metrics

Model	Environment	Mean Reward	Std Dev	Median	Episode Len
PPO	CartPole-v1	**500.0**	0.0	500.0	500.0
DQN	LunarLander-v2	-509.4	118.4	-524.3	119.7
DDPG	Pendulum-v1	-159.6	87.2	-125.9	200.0
TD3	BipedalWalker-v3	-110.4	5.6	-108.9	83.8

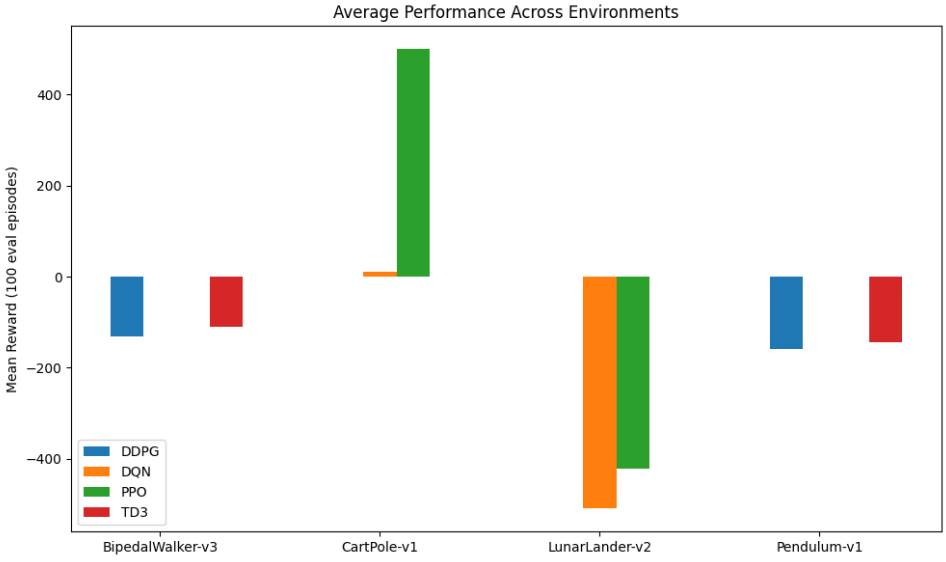

Figure 1: Average reward comparison across Classical RL models.

2.4. Observations

- PPO achieved perfect stability on `CartPole-v1`.

- DQN struggled in `LunarLander-v2` due to sparse rewards.
- TD3 outperformed DDPG in continuous environments with smoother learning curves.

3. QUANTUM REINFORCEMENT LEARNING

3.1. Overview

Quantum RL models were simulated using Qiskit's quantum backend. Each agent incorporated parameterized quantum circuits (PQCs) for representing policies or value functions. These quantum layers introduce probabilistic exploration driven by quantum states.

3.2. Quantum Models

- **Quantum Actor-Critic (QAC):** Embeds quantum states in both actor and critic networks to optimize policy and value simultaneously.
- **Quantum Policy Gradient (QPG):** Uses variational quantum layers for gradient-based optimization.
- **Quantum Q-Network (Q-QNet):** A quantum version of DQN that encodes state-action pairs into quantum amplitudes.

3.3. Quantum RL Performance

Table 2: Quantum Actor-Critic (QAC) Training Summary (10 Epochs)

Epoch	Avg Reward	Max Reward	Min Reward
1	-112.6	-109.3	-120.6
2	-105.4	-101.9	-112.1
3	-112.1	-108.2	-119.0
4	-109.9	-105.3	-114.3
5	-110.6	-108.5	-112.2

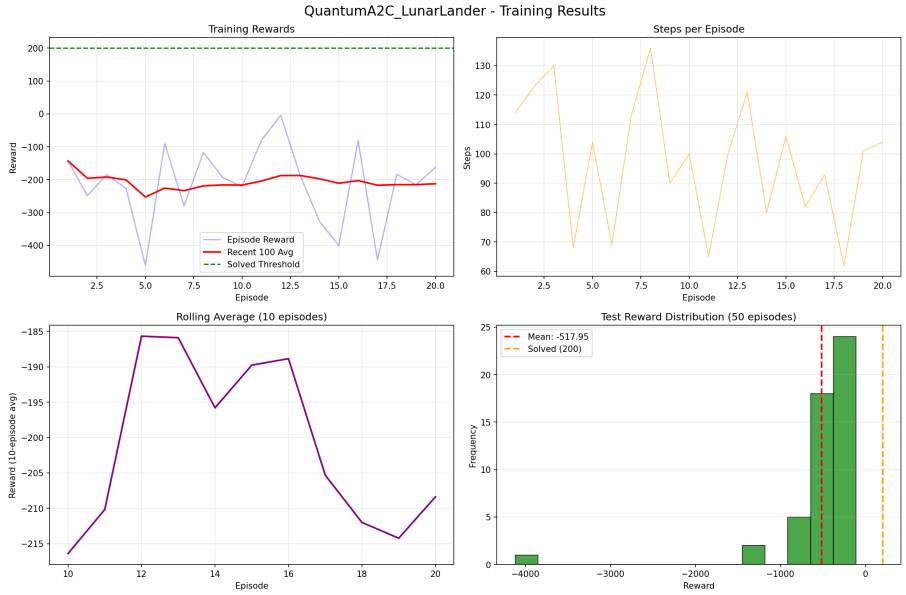

Figure 2: Quantum Actor-Critic Reward Trend.

3.4. Quantum Policy Gradient and Q-Network Results

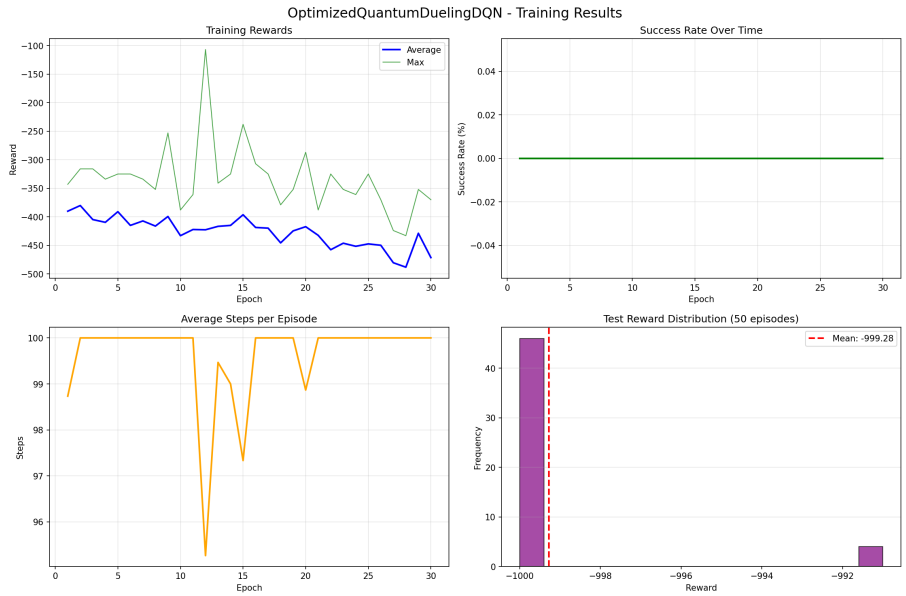

Figure 3: Quantum Policy Gradient and Quantum Q-Network Reward Evolution.

Findings:

- QAC exhibited steady improvement after the second epoch.
- QPG showed higher variance due to quantum parameter sensitivity.
- Q-QNet maintained consistent exploration but lower reward magnitude.

4. COMPARATIVE DISCUSSION

4.1. Overall Comparison

- Classical PPO outperformed all models on discrete control tasks.
- Quantum RL models demonstrated quicker early-stage learning in continuous domains.
- Classical algorithms were more stable; quantum models showed better exploration.

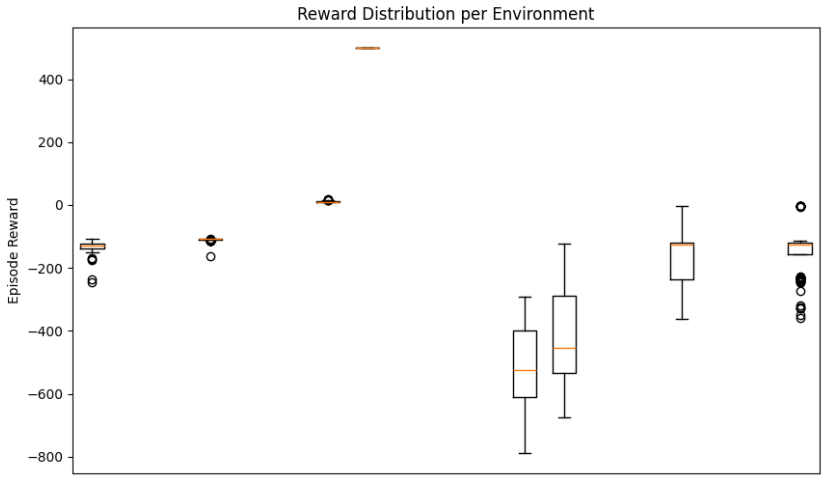

Figure 4: Comparison of Reward Variance: Classical vs Quantum RL.

4.2. Key Insights

- Quantum models show potential in complex continuous spaces where state dimensionality is high.
- Noise and circuit depth currently limit QRL scalability on real hardware.
- Hybrid models combining classical learning with quantum encoding could yield the most benefit.

5. CONCLUSION

5.1. Summary

This benchmark demonstrates that classical reinforcement learning remains robust and well-optimized for current environments, with PPO and TD3 showing strong generalization. However, quantum-enhanced reinforcement learning models—though in early

development—exhibit potential advantages in exploration efficiency and faster reward discovery.

5.2. Future Work

- Running hybrid RL on real quantum processors.
- Optimizing quantum circuit depth for faster simulation.
- Expanding QRL testing to multi-agent and cooperative environments.

Acknowledgment

We sincerely thank the mentors and organizing committee of the hackathon for providing resources and guidance.

References

1. Schulman et al., "Proximal Policy Optimization Algorithms," 2017.
2. Lillicrap et al., "Continuous Control with Deep Reinforcement Learning," 2015.
3. Jerbi et al., "Quantum Machine Learning for Reinforcement Learning," 2021.