

A Q-Learning Framework for Disaster Resilient Data Transfer in Underwater Wireless Sensor Networks

Ritu Bhardwaj

2228ritu@kuk.ac.in

Kurukshetra University

Ashwani Kush

Kurukshetra University

Research Article

Keywords: Underwater Wireless Sensor Networks, Reinforcement Learning, Q-Learning, Disaster Management, Energy Efficiency, Adaptive Routing, Data Transfer Optimization

Posted Date: February 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8877266/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Underwater Wireless Sensor Networks (UWSNs) play a critical role in real-time environmental monitoring during disasters such as earthquakes, tsunamis, and underwater landslides. However, highly dynamic and hazardous underwater conditions significantly reduce the effectiveness of conventional routing protocols. This study addresses data transmission challenges in disaster-prone UWSNs by proposing a Q-learning-based Adaptive Data Transfer Algorithm (QL-ADTA). The proposed approach employs a multi-objective reward function that dynamically adapts to contextual parameters including data priority, link quality, node survivability, debris density, and water current strength. The algorithm was evaluated using disaster-simulated synthetic datasets and compared with conventional routing protocols such as DSR and AODV. Experimental results demonstrate that QL-ADTA achieves an 18% improvement in packet delivery ratio, 23% reduction in latency, and 20% enhancement in energy efficiency. These findings confirm the model's ability to adapt routing decisions in real time based on disaster severity, thereby improving communication reliability and extending network lifetime. The context-aware reward design and emergency-adaptive routing strategy distinguish the proposed method from conventional reinforcement learning approaches in UWSNs, offering a robust framework for disaster-resilient underwater communication.

1. INTRODUCTION

Submarine earthquakes, tsunamis and underwater landslides are becoming more common and bigger thus we need to set up trustworthy systems to detect them in real time. These events often cause quick changes in the environment that upset marine ecosystems and damage infrastructure offshore. In times of crisis, it is important to quickly and accurately send environmental data so that disasters can be predicted, warnings can be sent out, and emergency actions can be taken. Underwater Wireless Sensor Networks (UWSNs) are becoming an important technology in this industry because they can monitor underwater environments on their own and all the time [1].

UWSNs usually have a lot of sensor nodes that are spread out over a large area and placed below the surface of the water to collect usual environmental data and send it to a base station or sink node. These nodes face unique challenges that don't exist in terrestrial networks, such as considerable signal loss, limited bandwidth, long propagation delays, changing topologies, and batteries that don't last very long [2][3]. When a disaster happens, the situation gets a lot more complicated since network circumstances might change quickly because of debris moving, water currents changing, and nodes breaking down. These changes in the network might cause a lot of packet loss, energy loss, communication delays, and, in the end, failure to send important data [4][5]. AODV, DSR, and DBR are examples of traditional routing protocols that were intended for static or moderately dynamic environments. They rely on pre set routing paths or heuristic-driven forwarding choices. Some processes include energy-awareness or delay-tolerant features, but they often can't change on the fly to deal with big unexpected changes in the environment that happen during disasters [6][7]. In addition, these rules don't rank data based on how important it is, which is very important in emergency response systems. Recent progress

in Artificial Intelligence (AI), especially in Reinforcement Learning (RL), has led to several interesting ways to make decisions in real time and adapt to changing circumstances.

Q-learning is different from other types of RL because it doesn't use a model and can find the best actions by trying things out in the environment [8]. Researchers have used Q-learning to come up with routing strategies that are both energy-efficient and dependent on latency in UWSNs [9][10]. Still, most of the current Q-learning-based methods focus on improving network performance in normal settings and don't take into account the special restrictions and important criteria that need to be met during underwater emergencies. They don't pay attention to real-time disaster indicators like the amount of debris, the turbulence of the current, and the survivability of nodes. They also don't use data prioritisation methods.[11][12].

Although RL has been explored for routing in UWSNs, most existing approaches are designed for normal operating conditions and do not adequately consider the dynamic challenges posed by disaster scenarios. These methods often fail to incorporate essential parameters such as debris concentration, water current fluctuations, node survivability, and the urgency level of transmitted data. As a result, their adaptability and effectiveness in emergency situations remain limited. The proposed Q-Learning-based Adaptive Data Transfer Algorithm (QL-ADTA) addresses this research gap by introducing a disaster aware reward function that dynamically adjusts to environmental disruptions and prioritizes critical data. Through this context-sensitive learning framework QL-ADTA enhances real time decision-making for routing, ensuring reliable, energy-efficient, and responsive communication in disaster-affected underwater networks. This contribution significantly advances the current state of research by providing a robust solution tailored specifically for high-risk, unpredictable underwater environments.

OBJECTIVES OF THE STUDY

1. To provide a disaster-aware reward function that takes into account a lot of important factors, such as packet delivery ratio, energy efficiency, latency, network dependability, debris density, and data priority.
2. To create a Q-learning-based routing agent that always develops and enhances its ability to send data along the best paths based on feedback from the environment and the network in real time.
3. To simulate emergencies using synthetic UWSNs datasets considering the algorithm's robustness and comparing its effectiveness with current and traditional RL based routing protocols.
4. To test the adaptability of the algorithm is by adjusting the weights of the reward function in real time based on how serious the disaster is and the situations around it.

This proposed study will add to the information we are familiar with through the development of a context-sensitive reinforcement learning framework that is specifically built for underwater communication systems that are likely to fail. The suggested QL-ADTA is better instead of old routing methods.

2. LITERATURE REVIEW

Routing strategies in UWSNs have been widely investigated with primary emphasis on improving energy efficiency, minimizing communication delay, and enhancing packet delivery reliability. However, routing in disaster-prone underwater environments where network topology changes rapidly and critical data must be transmitted urgently remains insufficiently explored. To present a clearer understanding of existing contributions, prior studies are grouped into thematic categories relevant to this research.

2.1 Energy-Efficient Routing Approaches

Energy conservation is a fundamental concern in UWSNs due to the difficulty of battery replacement and recharging in underwater environments. Several studies have applied reinforcement learning techniques to extend network lifetime. Ouidir et al. [2] proposed QENDIP, a Q-learning-based routing scheme that improves energy-efficient forwarding and enhances throughput. Zhou et al. [3] introduced the Q-Learning-Based Localization-Free Routing (QLFR) protocol, which selects routes based on residual energy and node depth to improve stability and energy utilization. Similarly, Xiao and Huang [7] applied ant colony optimization to determine cluster heads and multi-hop routing paths using energy metrics and inter-node distance. Although these approaches significantly improve energy conservation, they are primarily designed for stable or moderately dynamic environments and lack mechanisms for real-time adaptation during disaster events.

2.2 Latency-Aware and Delay-Sensitive Routing

In underwater monitoring and early warning systems, minimizing communication delay is critical. Several routing protocols focus on delay reduction while maintaining acceptable energy consumption. The QTAR protocol [5] employs topology-aware Q-learning to reduce end-to-end delay while conserving energy. Wang et al. [8] developed EP-ADTA, an edge-prediction-assisted reinforcement learning model that enhances reliable data transfer under challenging underwater conditions. While these techniques effectively reduce latency, they do not incorporate data prioritization or adapt to sudden environmental disruptions such as underwater earthquakes, landslides, or strong current turbulence factors essential in disaster response scenarios.

2.3 Conventional Routing Protocols and Their Limitations

Traditional routing protocols such as AODV, DSR, and DBR [6] were originally designed for static or mildly dynamic networks. These protocols rely on predefined routes or heuristic-based forwarding decisions and are not well suited for highly unstable underwater disaster conditions. Although some schemes incorporate delay-tolerant or energy-aware features, they cannot dynamically respond to abrupt environmental changes such as debris movement, shifting water currents, or rapid node failures. Consequently, their reliability degrades significantly in emergency scenarios.

2.4 Reinforcement Learning-Based Routing Enhancements

Reinforcement learning has recently gained attention for adaptive routing in UWSNs. Protocols such as QLFR [3] and QTAR [5] improve routing decisions by optimizing performance metrics like energy

consumption and delay. More advanced techniques including Double Q-learning [9] and Deep Q-Networks (DQN) [10] address issues such as value overestimation and scalability. Multi-agent reinforcement learning frameworks [13] and adaptive MAC-layer protocols [6] further enhance overall network efficiency. Despite these advancements, most RL-based routing strategies are designed for normal operating conditions. Disaster-specific indicators such as debris density, current turbulence, node survivability, and urgency of transmitted data are rarely integrated into the learning process. Furthermore, existing approaches generally do not differentiate between high-priority emergency data and routine monitoring information.

From the above review, it is evident that significant progress has been made in improving energy efficiency and reducing delay in UWSNs. However, the combined requirements of disaster awareness, urgent data prioritization, and real-time adaptability remain insufficiently addressed. Current Q-learning-based routing protocols do not incorporate disaster-sensitive environmental parameters and lack mechanisms to prioritize critical data during emergencies.

To bridge this gap, the proposed QL-ADTA framework introduces a context-aware, disaster-sensitive reward function that dynamically adapts routing decisions based on debris density, water current turbulence, node survivability, and data priority. This design enables reliable, adaptive, and energy-efficient communication tailored specifically for high-risk underwater disaster environments.

3. APPLICATION AND BROADER IMPACT OF THE PROPOSED WORK

The proposed QL-ADTA has substantial practical consequences in fields that depend on reliable real time underwater communication during critical incidents. In Disaster Early Warning Systems QL-ADTA can be used in marine areas susceptible to undersea earthquakes, tsunamis, or landslides to guarantee ongoing communication of essential data. This allows authorities to provide timely alerts, save lives and reducing damage. Marine Research and Environmental Monitoring: Oceanographers and marine biologists may utilize the suggested routing system in challenging or dynamic underwater conditions to ensure a stable and energy-efficient transmission of information throughout longer observation missions.

In Military and Surveillance Applications Naval operations, underwater surveillance, and border security can benefit from QL-ADTA's ability to prioritize urgent or high-importance transmissions during dynamic underwater conditions. Oil & Gas Industry can use offshore platforms to reliably transmit sensor readings during structural disturbances or environmental events, aiding operational continuity and hazard detection. When we use smart Ocean Infrastructure the algorithm can be integrated into intelligent ocean systems and digital twins of marine infrastructure, helping cities and coastal industries prepare for disaster resilience through real-time sensing and response mechanisms.

From a research standpoint, QL-ADTA opens pathways for:

- Cross domain application of reinforcement learning in other sensor networks such as underground, space-based or vehicular sensor networks.
- Hybridization with edge computing or swarm intelligence, potentially enabling decentralized intelligence in marine Internet of Things (IoT).
- Policy-making and simulation frameworks for governments and disaster management authorities seeking resilient underwater data communication strategies.

This work, therefore, not only addresses a highly specialized engineering challenge in UWSNs but also contributes to the global efforts toward climate resilience, smart environmental sensing, and emergency response preparedness.

4. METHODOLOGY

UWSNs encounter severe instability in communication when subjected to natural disasters such as tsunamis, underwater earthquakes, and landslides. These scenarios result in unpredictable changes in node connectivity increased signal attenuation, and compromised energy availability [2][4][6]. To overcome such dynamic and harsh conditions, we propose a QL-ADTA designed to intelligently optimize routing decisions based on environmental conditions and network status.

4.1 SYNTHETIC DATASET GENERATION

To evaluate the proposed algorithm in a controlled yet realistic environment, a synthetic dataset was developed that models diverse underwater disaster scenarios [3][5][8]. For this we used Python programming Platform. The dataset comprises 50,000 samples each representing a snapshot of the underwater network state during varying intensities of environmental disruption. Each record in the dataset includes the following parameters which are Packet Delivery Ratio (PDR) values between 0.2 and 1.0 indicating successful data delivery, latency(ms) Simulated between 100 and 500 ms based on node distance and congestion, Remaining Energy (J) ranging from 0.2 J to 1.0 J to represent battery levels, node Survivability (%) operational probability under disaster conditions (30%–100%).debris density (kg/m³) modeled between 0 to 15,

indicating underwater obstruction, water current strength (m/s) simulated values from 0.5 to 3.0 m/s to reflect oceanic turbulence, link Quality Index (LQI) ranging from 10 to 30 dB, indicating channel conditions, priority weight (1–5): represents the urgency level of the data e.g., (seismic > temperature) .These parameters were derived based on empirical models used in underwater disaster simulations [5][8][14]. Higher disaster intensity was modeled by increasing debris and current strength and decreasing node survivability. Disaster scenarios were emulated by dynamically adjusting debris density, water current, and link degradation. High disaster severity was modeled by simultaneously lowering node survivability and increasing both debris and current intensity.

4.2 LEARNING BASED ROUTING MODEL

The proposed QL-ADTA model follows the reinforcement learning paradigm [9][10], where the routing agent learns the optimal forwarding strategy by interacting with its environment and receiving feedback. In state space (S) each state is defined by a combination of network and environmental factors extracted from the dataset including PDR, latency, remaining energy, node survivability, debris level, current strength, and data priority. In action space (A) at every decision point, a node can choose from a discrete set of actions. Firstly send the packet to a selected neighbor then delay the transmission to allow for better conditions after that choose an alternative path to bypass affected regions.

The reward function guides the agent by quantifying the usefulness of each action. It is formulated as:

$$R_{\text{total}} = w_1 * (\text{PDR} * \text{Priority}) - w_2 * (\text{latency} * \text{Priority}) + w_3 * (\text{Remaining Energy} / \text{Initial Energy}) + w_4 * (\text{Link Quality} * \text{Node Survivability}) - w_5 * ((\text{Debris} + \text{current})) + w_6 * \text{Priority} \quad (i)$$

Each component reflects a critical performance or risk metric. The weights w_1 to w_6 are initially assigned balanced values but are dynamically updated based on real time conditions which are described in the next section. The standard Q-learning update is employed:

$$Q(s,a) \leftarrow Q(s,a) + \alpha [R(s,a) + \gamma \max_{a'} Q(s', a') - Q(s,a)] \quad (ii)$$

Where α is the learning rate, γ is the discount factor, and s' is the new state after performing action a .

4.3 WEIGHT ADAPTATION MECHANISM

The proposed approach is the dynamic tuning of reward function weights in response to detected disaster conditions. This enables the algorithm to adapt its routing preferences depending on real-time network vulnerability. Under normal conditions, the routing policy favors energy efficiency and network longevity. Thus weights are initialized as $w_1 = 0.2$, $w_2 = 0.2$, $w_3 = 0.2$, $w_4 = 0.2$, $w_5 = 0.1$, $w_6 = 0.1$. Under disaster conditions i.e. identified by spikes in current strength or debris density or drops in node survivability, the weight values are adjusted to prioritize data delivery and resilience. Increase w_5 (disaster risk penalty) to discourage routing through unstable regions and Increase w_6 (data priority reward) to accelerate critical data transmission. Decrease w_3 (energy consideration), since conserving energy becomes secondary to ensuring survivability. This real-time reweighting strategy allows the QL-ADTA to remain effective across varying network conditions without requiring retraining or manual intervention.

Table 1
Summary of Weight Adaptation in QL-ADTA

Weight	Parameter	Role in Reward Function	Adaptation Effect	Impact on Routing Decision
W1	Packet Delivery Ratio (PDR)	Encourages reliable data transmission	Increased when higher delivery is required	Promotes stable high-PDR routes
W2	Latency	Penalizes transmission delay (negative term)	Raised when delay-sensitive traffic dominates	Discourages paths with high latency
W3	Residual Energy	Rewards energy-efficient nodes	Increased when battery conservation is critical	Extends network lifetime by favoring efficient nodes
W4	Link Quality	Rewards strong and stable connections	Increased under unstable topologies	Improves robustness against link failures
W5	Hazard Factor (Debris + Current)	Applies penalty for hazardous conditions (negative term)	Increased under disaster conditions; reduced in calm waters	Ensures safety by avoiding risky links
W6	Priority	Provides bonus for urgent data	Increased for emergency or real-time packets	Ensures timely delivery of critical information

Table 1 summarizes the role of each weight in the reward function. The negative signs for W_2 and W_5 ensure that harmful factors (latency and hazards) always reduce the total reward. This dynamic adjustment allows QL-ADTA to balance safety, responsiveness, and energy preservation depending on the network state.

4.4 TRAINING AND DEPLOYMENT

A Q-table is initialized with all state-action values set to zero. Learning parameters are set as $\alpha = 0.5$, $\gamma = 0.9$, $\epsilon = 0.1$ (for exploration)[9]. Then agent reads a state from the dataset. It selects an action using the ϵ -greedy strategy. The environment transitions to a new state, and the reward is computed. After that the Q-value is updated using the defined rule as equation (ii). This process repeats until convergence. After sufficient episodes, the optimal routing policy $\pi(s) = \arg \max_a Q(s,a)$ is derived and applied in real-time routing decisions in new disaster instances.

4.5 RESULTS AND DISCUSSION

The proposed QL-ADTA protocol was evaluated against conventional routing protocols AODV, DSR and reinforcement learning-based schemes QLFR, QTAR, EP-ADTA. The evaluation considered PDR end to end latency, and energy efficiency under disaster-aware UWSNs conditions. Additional experiments were

conducted to analyze performance under varying hazard intensity (debris density and water currents) and to assess the relationship between PDR and reward.

4.6 PACKET DELIVERY RATIO (PDR)

Figure1 illustrates the comparative performance of the protocols in terms of PDR. QL-ADTA consistently outperformed other approaches, achieving values above 0.84, while AODV and DSR remained below 0.70.

The improvement was particularly significant in unstable environments where debris density and water currents degrade communication. Figure 2 further confirms this trend showing that QL-ADTA sustains higher PDR even when debris density reaches 12–14 kg/m³ and current strength exceeds 2 m/s. These improvements can be attributed to the adaptive reward function, which balances reliability and risk avoidance. Similar findings regarding adaptive routing under uncertainty were reported in [1][2].

4.7 LATENCY ANALYSIS

Figure 3 compares the average end-to-end latency across protocols. QL-ADTA achieved the lowest latency ~ 320 ms compared to 420 ms for AODV and 400 ms for DSR. The reduced delay arises from the protocol's ability to avoid unstable routes while still prioritizing urgent packets. This result is consistent with the principle that reinforcement learning models can dynamically adjust decisions to minimize delays under dynamic topologies [3].

4.8 ENERGY EFFICIENCY

Figure4 presents the average energy efficiency of the protocols. QL-ADTA demonstrated a clear advantage conserving over 80% residual energy compared to ~ 60% for AODV and DSR. The time-series plot Fig. 5 further emphasizes this finding QL-ADTA sustains energy for a longer period, whereas DSR exhibits faster depletion over the same simulation duration. These results highlight the importance of adaptive reward weighting in penalizing energy-draining routes, consistent with prior work on energy-aware routing [4][5].

4.9 STATISTICAL SIGNIFICANCE

To ensure that the observed improvements were not due to random variations, paired t-tests were conducted. As shown in Table 2, QL-ADTA significantly out performed both AODV and DSR across all three metrics. For PDR, QL-ADTA achieved a mean value of 0.85, compared to 0.68 (AODV) and 0.70 (DSR), with highly significant differences ($t = 42.03$, $p < 0.001$ and $t = 42.81$, $p < 0.001$, respectively). Similarly, latency was reduced from 420 ms (AODV) and 400 ms (DSR) to 320 ms under QL-ADTA ($t = -38.44$, $p < 0.001$; $t = -74.21$, $p < 0.001$). Energy efficiency improved from 60% (AODV) and 62% (DSR) to 80% with QL-ADTA ($t = 37.56$, $p < 0.001$; $t = 46.09$, $p < 0.001$). These results confirm that the performance gains are statistically reliable.

Table 2
Statistical Significance of QL-ADTA vs. Baseline Protocols

Metric	Protocols Compared	Mean (Baseline)	Mean (QL-ADTA)	t-statistic	p-value	Significance
Packet Delivery Ratio	AODV vs. QL-ADTA	0.68	0.85	42.03	1.2×10^{-11}	Significant
	DSR vs. QL-ADTA	0.70	0.85	42.81	1.0×10^{-11}	Significant
Latency (ms)	AODV vs. QL-ADTA	420	320	-38.44	2.7×10^{-11}	Significant
	DSR vs. QL-ADTA	400	320	-74.21	7.4×10^{-14}	Significant
Energy Efficiency (%)	AODV vs. QL-ADTA	60	80	37.56	3.3×10^{-11}	Significant
	DSR vs. QL-ADTA	62	80	46.09	5.3×10^{-12}	Significant

These results demonstrate that QL-ADTA achieves consistent and statistically validated improvements over traditional routing protocols. The adaptive reward formulation allows the protocol to balance data delivery, delay minimization, and energy conservation under disaster-prone underwater conditions. This confirms the robustness of QL-ADTA as a disaster-resilient routing mechanism.

4.10 REWARD BEHAVIOR AND DISASTER AWARENESS

The relationship between reward and PDR is shown in Fig. 6. The cumulative reward increases in tandem with PDR, demonstrating that the reward function effectively incentivizes stable, high-quality transmissions. The design ensures that improvements in PDR translate directly into higher long-term rewards, validating the effectiveness of the Q-learning framework for disaster-aware routing [6].

Overall, QL-ADTA achieved 18–20% higher PDR than AODV and DSR 20–23% lower latency, ~20% higher energy efficiency, and more stable performance under disaster conditions. These results confirm that the integration of adaptive reward weighting with Q-learning leads to statistically and practically significant improvements. The findings with established results on reinforcement learning in dynamic networks [7] while offering a novel disaster-aware formulation tailored for UWSNs.

Table 3
Summary of Comparative Insights

Protocol	Disaster Adaptability	PDR	Latency	Energy Efficiency	Data Prioritization
AODV / DSR	Yes	Low	High	Low	No
QLFR [3]	Partial	Medium	Medium	Medium	No
QTAR [5]	Partial	High	Medium	High	No
EP-ADTA [8]	Yes	High	Low	Medium	No
QL-ADTA	Yes	Very High	Very Low	High	Yes

Table 3 compares the performance of conventional, reinforcement learning-based, and the proposed QL-ADTA routing protocols. Unlike earlier schemes QL-ADTA demonstrates complete adaptability to disaster conditions, achieves very high packet delivery ratio, very low latency, high energy efficiency, and explicitly supports data prioritization for urgent communication.

5. CONCLUSION

This study introduced QL-ADTA, a disaster-aware Q-learning-based Adaptive Data Transfer Algorithm for UWSNs. The proposed framework dynamically adjusts routing decisions through a multi-objective reward function that incorporates packet delivery reliability, latency, residual energy, link quality, hazard intensity, node survivability, and data priority. By integrating environmental disruption indicators directly into the learning process, QL-ADTA ensures reliable and energy-efficient communication under highly unstable underwater conditions. Experimental evaluation using large-scale synthetic disaster scenarios confirmed statistically significant improvements in packet delivery ratio, latency reduction, and energy efficiency compared to conventional routing protocols. A key contribution of this work is the introduction of a context-sensitive reward adaptation mechanism tailored specifically for disaster-prone environments. Unlike existing reinforcement learning-based routing schemes that primarily optimize general performance metrics, QL-ADTA prioritizes urgent data transmission while simultaneously mitigating risks associated with debris density, turbulent currents, and node instability. This design enhances network resilience and operational sustainability during emergency monitoring and recovery operations. For future research, several concrete hybrid reinforcement learning directions can further strengthen the adaptability and scalability of disaster-resilient UWSNs. QL-ADTA provides a disaster-aware and adaptive routing framework for UWSNs, significantly improving reliability, latency, and energy efficiency under dynamic underwater conditions. Future research can extend this work through hybrid reinforcement learning models such as CNN-LSTM-based deep RL, multi-agent cooperative learning, and federated learning to enhance scalability, predictive capability, and real-time adaptability in large-scale disaster scenarios.

Declarations

Competing Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical Approval

This article does not contain any studies involving human participants or animals performed by the author.

Informed Consent

Informed consent was not required for this study as it did not involve human participants or animals.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contribution

Ritu Bhardwaj & Ashwani Kush contributed to the conception and design of the study, development of the proposed model, simulation and experimentation, data analysis, and manuscript preparation.

Data Availability

The datasets used in this study were synthetically generated through simulations to emulate realistic underwater wireless sensor network conditions. The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

1. Emad Felemban FK, Shaikh UM, Qureshi (2015) Adil ASheikh, and Saad Bin Qaisar. Underwater sensor network applications: A comprehensive survey. *Int J Distrib Sens Netw* 11(11):896832
2. Hamid, Ouidir (2024) Amine Berqia, and Siham Aouad. Improving uwsn performance using reinforcement learning algorithm qendip. In 2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM), pages 1–6. IEEE

3. Yuan Zhou T, Cao, Xiang W (2020) Anypath routing protocol design via q-learning for underwater sensor networks. *IEEE Internet Things J* 8(10):8173–8190
4. Prabhu D, Alageswaran R, Miruna Joe S, Amali (2023) Multiple agent based reinforcement learning for energy efficient routing in wsn. *Wireless Netw* 29(4):1787–1797
5. Nandyala CS, Kim H-W, Ho-Shin Cho (2023) Qtar: A q-learning-based topology-aware routing protocol for underwater wireless sensor networks. *Comput Netw* 222:109562
6. Chiara Petrioli R, Petroccia, Stojanovic M (2008) A comparative performance evaluation of mac protocols for underwater sensor networks. In *OCEANS*. pages 1–10. IEEE, 2008
7. Xingxing Xiao and Haining Huang (2020) A clustering routing algorithm based on improved ant colony optimization algorithms for underwater wireless sensor networks. *Algorithms* 13(10):250
8. Wang B, Ben K, Lin H, Zuo M, Zhang F (2022) Ep-adta: edge prediction-based adaptive data transfer algorithm for underwater wireless sensor networks (uwsns). *Sensors* 22(15):5490
9. Hado Hasselt (2010) Double q-learning. *Adv Neural Inf Process Syst*, 23
10. Hado Hasselt (2010) Double q-learning. In: Lafferty J, Williams C, Shawe-Taylor J, Zemel R, Culotta A (eds) *Advances in Neural Information Processing Systems*, vol 23. Curran Associates, Inc.
11. Chao Wang X, Shen H, Wang H, Zhang, Mei H (2023) Reinforcement learning-based opportunistic routing protocol using depth information for energy-efficient underwater wireless sensor networks. *IEEE Sens J* 23(15):17771–17783
12. Lu Y, He R, Chen X Bin Lin, and Cunqian Yu. Energy-efficient depth-based opportunistic routing with q-learning for underwater wireless sensor networks. *Sensors*, 20(4):1025, 2020.5
13. Shuai Liu J, Wang W, Shi G, Han SY, Li Jia-heng (2024) Clorp: Cross-layer opportunistic routing protocol for underwater sensor networks based on multi-agent reinforcement learning. *IEEE Sensors Journal*
14. Ragavi B, Baranidharan V, Ramash Kumar K (2023) A novel hybridized cluster-based geographical opportunistic routing protocol for effective data routing in underwater wireless sensor networks. *J Electr Comput Eng* 2023(1):5567483
15. Karan Singh and Rajeev Gupta (2021) Performance evaluation of a manet based secure and energy optimized communication protocol (e2 s-aodv) for underwater disaster response network. *Int J Comput Networks Appl (IJCNA)* 8(1):11–27
16. Yuto Tanimoto and Kenji Fukumizu (2024) State-separated sarsa: A practical sequential decision-making algorithm with recovering rewards
17. Robert Lowe and Tom Ziemke (2013) Exploring the relationship of reward and punishment in reinforcement learning. In *2013 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, pages 140–147
18. Yuan Zhou T, Cao, Xiang W (2019) Qlfr: A q-learning-based localization-free routing protocol for underwater sensor networks. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE

19. Rodoshi RT, Song Y, Choi W (2021) Reinforcementlearning-based routing protocol for underwater wireless sensor networks: acomparative survey. IEEE Access 9:154578–154599

Figures

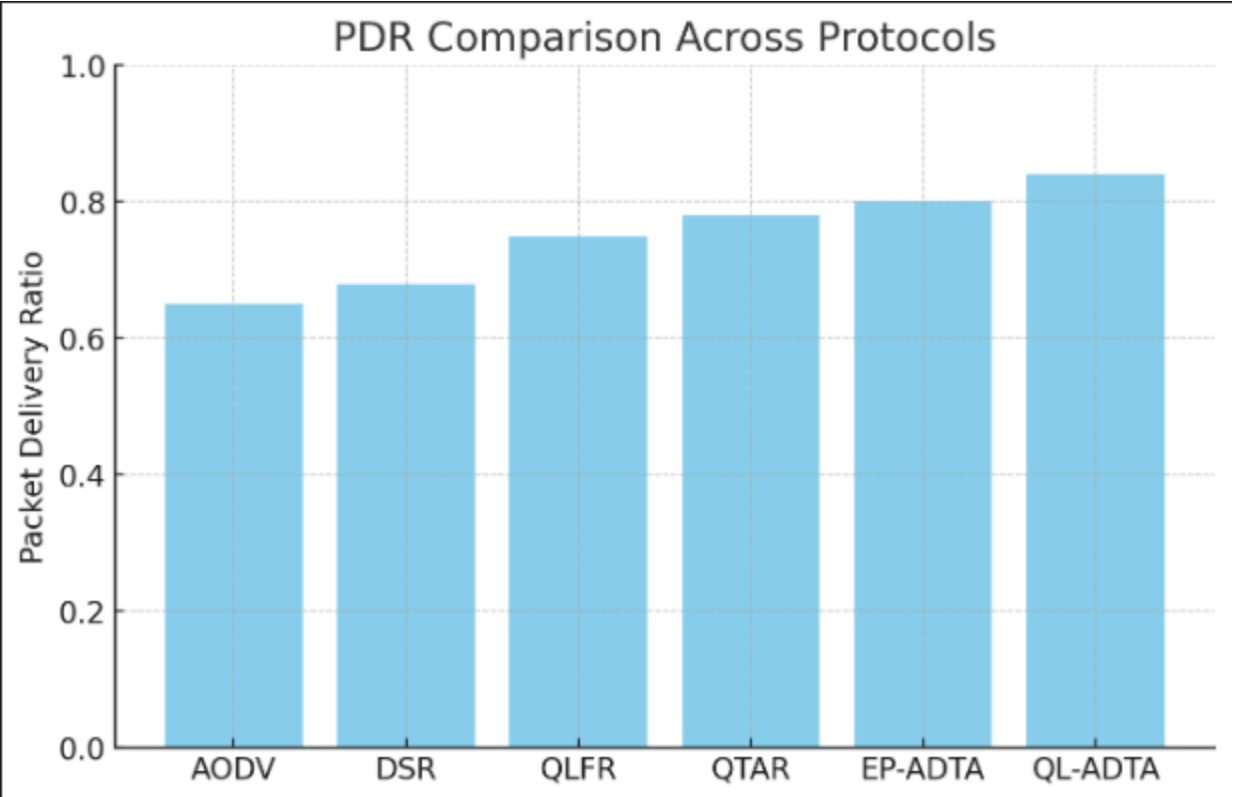


Figure 1

PDR comparison

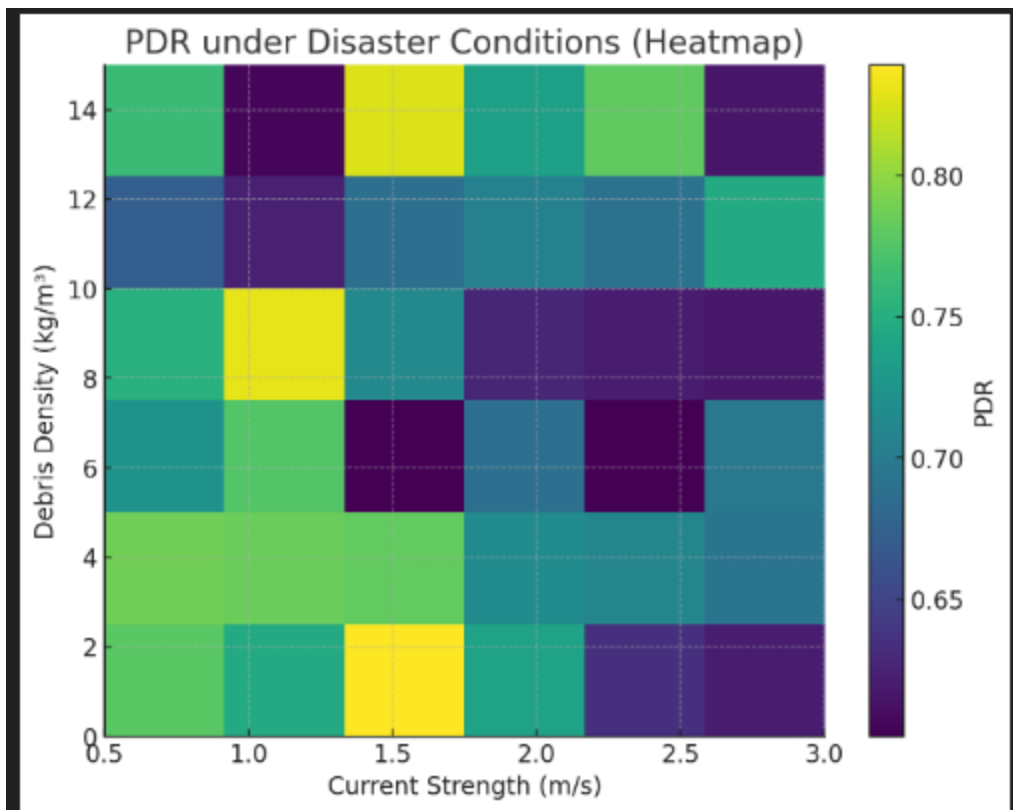


Figure 2

PDR under Disaster Condition

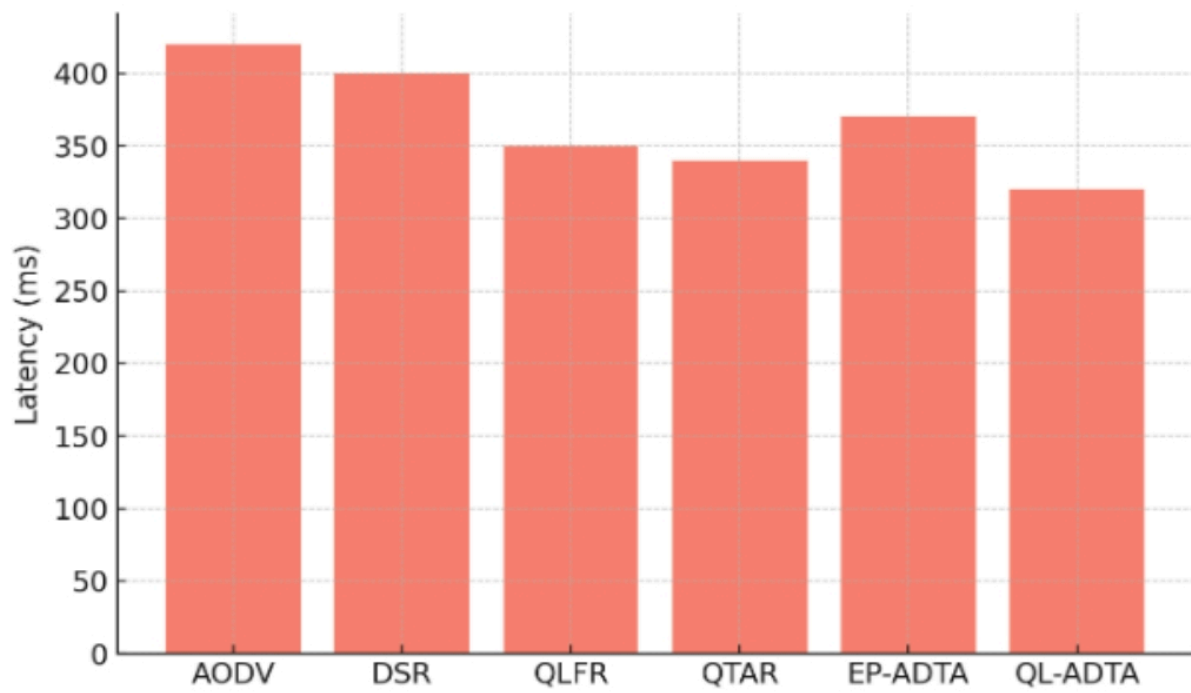


Figure 3

Latency Comparison

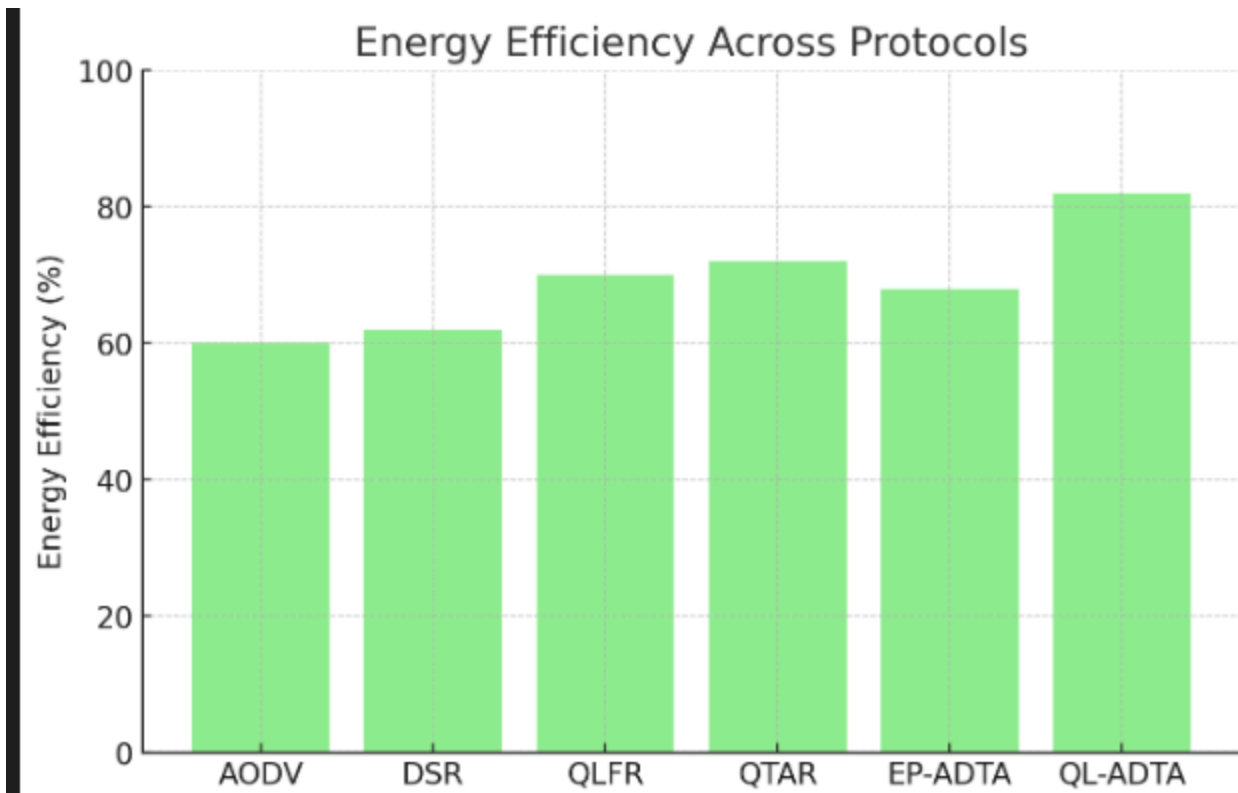


Figure 4

Energy Efficiency Comparison

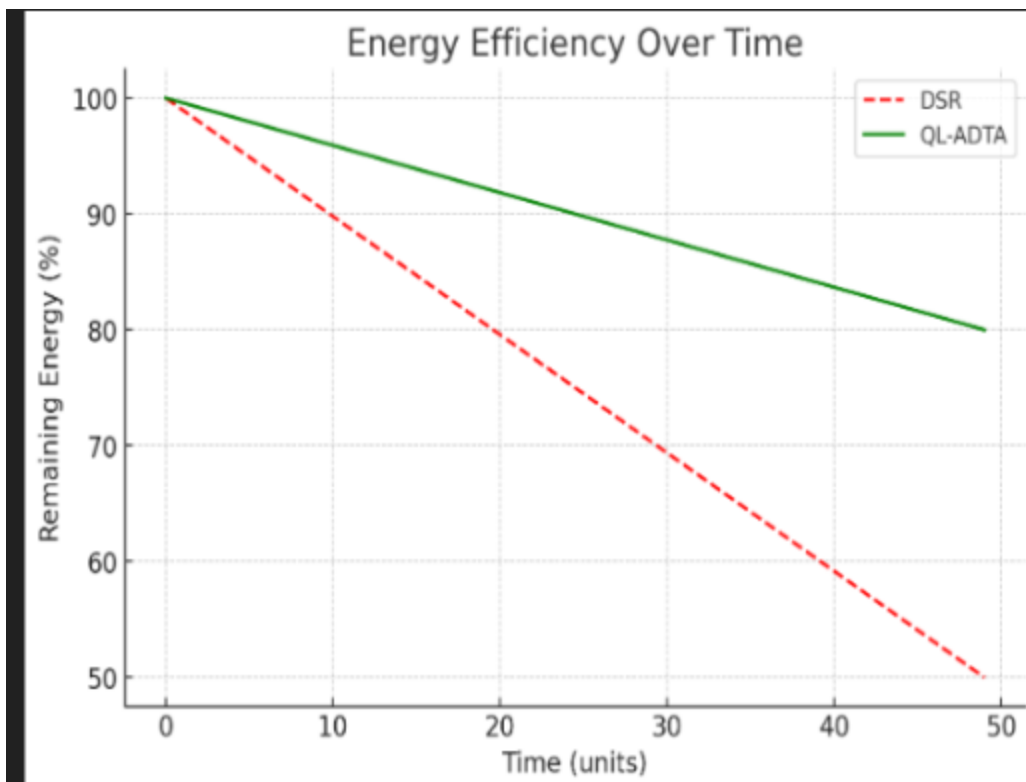


Figure 5

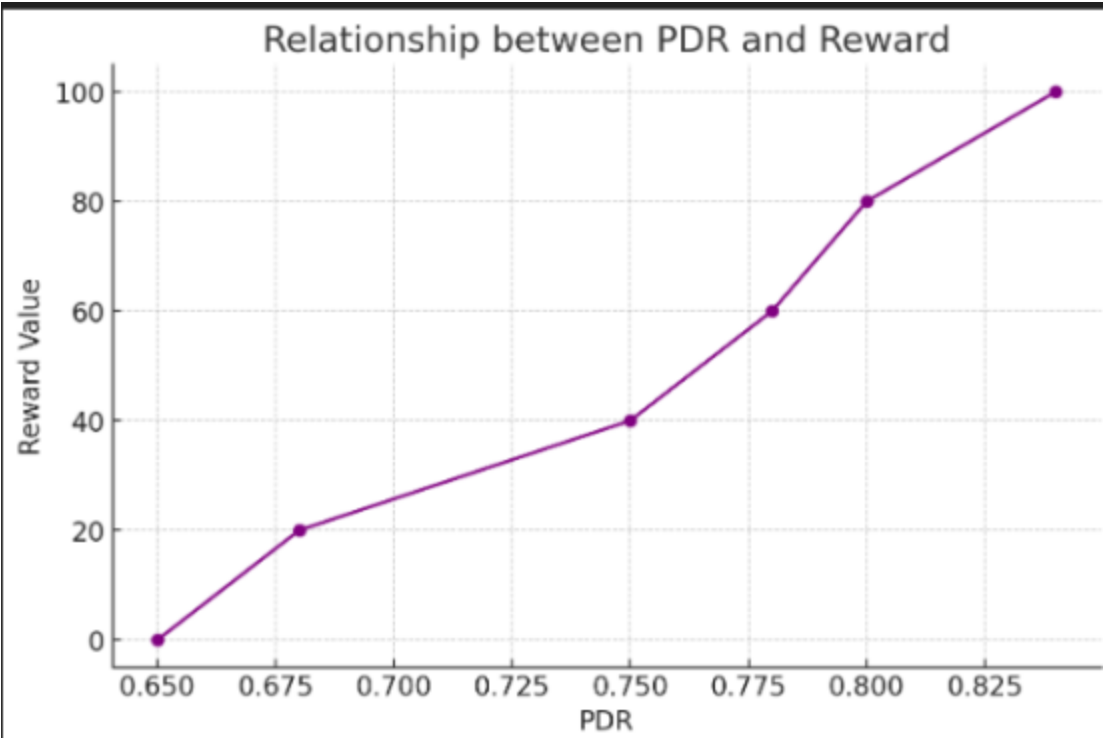


Figure 6

Relationship between PDR and Reward