Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DRL-Based Adaptive Routing for Energy-Efficient and Reliable Underwater Wireless Sensor Networks

Zeeshan Rasheed
DHA Suffa University
Shahzad Ashraf
taepyeong.info@gmail.com

Gachon University
Muhammad Junaid Rabbani
National University of Computer and Emerging Sciences

Article

Keywords: (UWSN) Under Water Sensor Network, (DRL) Deep Reinforcement Learning, (PDR) Packet Delivery Ratio, Network lifetime, (SNR) Signal Noise Ratio, Residual Energy

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11325854/v1

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

Additional Declarations: No competing interests reported.

DRL-Based Adaptive Routing for Energy-Efficient and Reliable Underwater Wireless Sensor Networks

Zeeshan Rasheed[1], Shahzad Ashraf[2,*], Muhammad Junaid Rabbani[3]

[1]Department of Computer Science, DHA Suffa University Karachi Pakistan; zeeshanrasheed1992@yahoo.com
[2,*]Department of Computer Engineering, Gachon University Seongnam South Korea; taepyeong.info@gmail.com
[3]Department of Electrical Engineering, National University of Computer and Emerging Sciences, Karachi, 75030, Pakistan; junaid.rabbani@nu.edu.pk
Corresponding Author: Shahzad Ashraf; taepyeong.info@gmail.com

Abstract

The amount of energy lost in underwater wireless sensor networks is greatly influenced by the choice of relay: a node that is frequently chosen might end up with an empty battery, and a weak acoustic link can cause retransmissions to be needed. Routing has to take into account both the conservation of relay energy and the need to make progress, ensure timely forwarding, and keep a viable next node available. The present study focuses on developing an Adaptive Routing method based on Deep Reinforcement Learning for energy efficient and reliable Underwater Wireless Sensor Networks. Prior to carrying out each transmission, the current node deletes from consideration any of its neighbors, that have insufficient residual energy, poor link quality, or fail to make any positive contribution to depth advancement towards the surface sink. The DRL policy then assesses the remaining possible relays according to their residual energy, SNR, depth advancement, forwarding delay, and void risk. Feedback on delivery updates the action values, and both link failures and voids are penalized. Testing in NS-3, using Aqua-Sim included nine baseline protocols and five different deployment scenarios. In a number of repeated trials involving 100 nodes each, the proposed approach obtained a mean PDR of 93.26%, a mean residual energy of 87.732%, and a mean delay of 0.982 s. ADRP-DQL achieved a PDR of 85.70% and a delay of 1.162 s, while Dynamic DRL maintained 81.029% residual energy. The proposed method was able to keep the network running for 662.6 rounds, as opposed to 625 rounds for Dynamic DRL. When adaptive reward balancing was not used, the network lifetime dropped to 560 rounds.

Keywords

(UWSN) Under Water Sensor Network, (DRL) Deep Reinforcement Learning, (PDR) Packet Delivery Ratio, Network lifetime, (SNR) Signal Noise Ratio, Residual Energy

1. Introduction

Underwater wireless sensor networks (UWSNs) rely primarily on acoustic communication because radio signals quickly weaken in water. These types of networks have been used to perform many different functions including; monitoring environmental conditions, tracking pollution levels, issuing disaster alerts, and inspecting underwater structures. Although there are some major limitations that affect the performance of UWSNs, such as limited bandwidth which limits data rate of transmission, long propagation delay that affects real-time communications, multipath effect that distorts signals, and Doppler shift caused by water movement that degrades link quality [1], overcoming these limitations will be necessary to ensure the successful deployment and use of UWSNs. Therefore, the routing function in UWSNs should provide reliable data transmission while preserving the limited battery life of the sensor nodes.

Node battery usage is determined by the selection of relay.Residual energy shows whether a node can continue forwarding packets, while link quality indicates the chance of successful transmission. Depth advancement measures progress toward the surface sink, forwarding delay affects timely delivery, and routing void risk shows whether a selected relay may fail to find the next hop. A decision based only on depth can choose a weak or depleted relay, while a decision based only on energy may select a node with

poor progress or an unreliable link. Energy efficient routing therefore needs a multi factor relay assessment rather than dependence on a single routing attribute.

Conventional protocols apply different rules to this forwarding decision. Depth Based Routing (DBR) forwards through shallower nodes without requiring complete location information, but depth alone ignores residual energy and link reliability and may create duplicate forwarding [2]. H2DAB uses dynamic addresses, which may lead to higher control expenses and potential delays in system operations [1]. Focused Beam Routing restricts relay discovery to specific geographic areas and dynamically adjusts transmission power according to the relay's location[3]. HHVBF modifies the direction of data flow (vectors) each time a node processes the packet. This provides greater flexibility in adapting to dynamic changes in the network, however, there can be duplicate data sent if the network topology is not uniform [4]. When network environments change unexpectedly, fixed routing protocols can reduce a node's lifetime. GEDAR addresses this issue with its ability to adapt dynamically to the environment by combining geographic routing, opportunistic routing, and the ability to adjust the depth of nodes located around void regions; thus providing more efficient and resilient communication in different types of environments [1]. Void-aware routing in underwater wireless sensor networks involves methods like switching to backup systems, choosing other relay nodes, and using information about the network's structure. The main problem is selecting relays that stop future gaps and also use energy in the best way possible. Good relay selection needs to keep the communication links strong, keep the transmission delays as low as possible, and make sure the route stays steady. By using knowledge about empty spaces and energy-saving techniques for relays, the network works better because there are fewer empty spaces and less energy is used.

Reinforcement learning is important in relay selection because picking the right relay doesn't just help with the current signal but also influences how the network performs in the future. A relay that is connected with a lot of current might drain too much battery power from a key forwarder, which could lead to issues later such as slow data movement or sending the packet to a less effective area for forwarding. Deep Q-Learning (DQL) helps figure out the long-term value of different relay actions by learning from experience over time, which means you don't have to make a separate entry for every possible situation and action. In the proposed study, Distributed Q-Learning (DQL) is used to help make decisions about how data is routed.This is necessary because the process of forwarding data is complicated and involves several factors such as changes in energy levels, the quality of the acoustic links, the positions of the nodes in space, and which nearby nodes are available. Recent advances in learning-based methods have made Underwater Wireless Sensor Networks (UWSNs) more adaptable; however, the existing designs do not completely align with the energy-focused approach explored in this research. The Adaptive Distributed Routing Protocol using DQL, or ADRP-DQL, uses a distributed method for DQL to help with smart routing in underwater settings. On the other hand, the Dynamic Reinforcement Opportunistic Routing (DROR) protocol uses reinforcement learning along with opportunistic forwarding and void recovery methods to improve routing efficiency even more [6]. QTAR uses Q-learning along with an understanding of the network's structure and identifies key points that, if removed, would split the network into separate parts.This helps avoid making forwarding decisions that might break the network's connectivity [7]. Alternative research assesses connection duration or amalgamates mobility, communication, and energy measurements [8], [9]. These techniques take into account valuable routing elements, and the suggested research investigates an energy-focused DRL framework that concurrently evaluates residual energy, SNR-based link quality, depth progression, forwarding latency, expected void risk, and adaptive reward priorities.

Therefore, increasing node battery life without compromising dependable and timely packet delivery is the research challenge. In UWSNs, choosing only the node with the highest residual energy might cause packets to go toward poor forward connection or increase latency, while continuously choosing the strongest link can deplete a crucial node. If rapid depth advancement results in a routing void, it may potentially become detrimental. The suggested study explores the possibility of balancing void risk, forwarding latency, depth advancement, residual energy, and connection quality in a single relay choice. How to ensure delivery dependability and future forwarding connection while conserving node energy is an unsolved technological problem.

Each potential relay is represented by the suggested DRL architecture using residual energy, forwarding delay, depth advancement, signal to noise ratio (SNR), and projected void risk. While SNR shows the present link quality, residual energy prioritizes nodes that can forward without hastening battery depletion. Latency deters sluggish forwarding channels, depth promotes advancement toward the sink, and void risk

assesses whether further transmission will be feasible. The reward function integrates these components, prioritizing network longevity and energy conservation. As network conditions evolve, an adaptive reward balance mechanism adjusts the influence of these parameters. This approach distinguishes the proposed method from DRL routing solutions that rely on fixed parameter priorities.

Learning-based routing in UWSNs presents several practical challenges. Feedback is delayed due to the propagation time of acoustic signals, and node mobility can render neighborhood information outdated before subsequent routing decisions. Policy changes at one node may influence the learning experiences of other nodes operating independently [10]. These factors can slow the learning process and result in suboptimal decision-making. Therefore, this study evaluates the proposed policy using empirical network results under modeled conditions, without assuming automatic convergence or optimal routing outcomes.

The framework is evaluated using NS-3 and Aqua-Sim. Comparative analysis includes DBR, H2DAB, GEDAR, FBR, and HHVBF as traditional and opportunistic methods, as well as ADRP-DQL, Dynamic DRL, DROR, and QTAR as learning-based approaches. Testing scenarios encompass standard operation, sparse deployment, partial node failures, random node placement, and high mobility. These scenarios demonstrate the method's ability to conserve energy resources under conditions of limited neighbor availability, dynamic network topology, and disrupted connectivity. The methods section details protocol configurations, mobility modeling, acoustic parameters, and simulation setup. Primary evaluation metrics include packet delivery ratio (PDR), end-to-end delay, residual energy, energy consumption per delivered packet, network lifetime, routing overhead, void packet failure, and routing decision time. These metrics assess the feasibility of reliable delivery with minimal energy and computational overhead. Residual energy, per-packet energy consumption, and network lifetime are critical for evaluating battery preservation, while PDR and latency reflect communication effectiveness. Overhead, void failure, and decision time quantify the operational cost of achieving high performance.

The core contribution of this study is a novel DRL-based relay selection method designed to optimize energy efficiency, extend node battery life, and enhance overall network longevity in UWSNs. The formulation incorporates residual energy, signal quality as measured by SNR, network depth, data forwarding delay, and the risk of routing voids to ensure that relay selection minimizes unnecessary transmissions and prevents premature depletion of critical nodes.It also has adaptive reward balancing, which adjusts routing priorities as energy scarcity, weak links, delay, or connectivity risk become more important. The entire system is compared to established conventional, opportunistic, topology aware, and learning-based routing algorithms in five scenarios. Component wise ablation removes the energy, link quality, delay, void penalty, and adaptive reward aspects one by one to determine their impact on energy preservation, PDR, delay, and future connectivity.

2. Literature Review

Routing in underwater wireless sensor networks (UWSNs) is shaped by the physical limits of the acoustic channel and by the restricted battery capacity of underwater nodes. Since these nodes are difficult to recharge or replace after deployment, an unsuitable forwarding decision can shorten network lifetime even when packet delivery appears acceptable at the current hop. Recent UWSN studies continue to report energy consumption, PDR, end to end delay, link reliability, routing voids, and topology changes as closely related routing concerns [2], [1], [11], [12]. The review below follows the pattern of the supervisor recommended paper. Each protocol is first discussed through its routing mechanism, and the limitation is then linked directly with the proposed study, whose main aim is to improve node battery life through energy aware DRL based relay selection.

2.1 Conventional and Opportunistic Routing Protocols

Depth Based Routing (DBR) was originally proposed as a location free routing protocol in which packets are forwarded through nodes with lower depth values, allowing data to move gradually toward the surface sink [2]. Its value lies in simplicity, because it avoids the need for full geographic localization, which is often costly and unreliable underwater. Recent routing studies still use DBR as a baseline because it represents a basic depth progress strategy for UWSNs [13], [14]. However, DBR mainly depends on depth and does not evaluate residual energy, SNR based link quality, forwarding delay, or routing void risk together. A node may therefore be selected repeatedly because of favorable depth, even when its battery is weak or its acoustic link is unstable. This practice can create energy gaps and reduce the network's lifespan, making DBR not useful for choosing relays that save battery life.

Hop by Hop Dynamic Addressing Based Routing (H2DAB) was created to help send data without needing full location details by giving surface sinks dynamic addresses based on the number of hops. [1]. This method pacts underwater nodes keep track of the right direction to send data even when it's difficult to know their exact location. However, keeping track of addresses requires extra work when nodes move or the network structure changes. H2DAB doesn't take into account things like leftover energy, how good the connection is, how much the network moves forward, delays, and the risk of gaps when choosing which relays to use. The protocol can keep track of the route direction even though it uses more energy because it sends updates or uses relays that aren't the best for keeping the network running for a long time. In a study that focuses on energy efficiency, this limitation is significant since the main problem is not merely whether a packet is transmitted upwards, but whether the relay chosen can send the packet reliably without its battery being used up quickly.

The FBR method makes use of location information to reduce the area in which relays may be located, so that the search is concentrated between the origin point and the destination[3]. By forwarding the data directly to the target, it avoids unnecessary flooding and enhances the transmission power whenever there is no relay at a lower level. This means that FBR is easier to manage than forwarding based on unlimited broadcasting. However, it should be noted that determining the optimal path and adjusting the power levels can be expensive when there are few devices in the area. Using excessive power to retransmit signals repeatedly can cause the battery to be drained quickly, particularly in a region where no relay devices are available. Furthermore, FBR does not take into account residual energy, the connection quality according to the signal-to-noise ratio, forwarding delay, progress in depth, or the risk of forming voids. Therefore, although it is moving in the right direction and is improving, it still does not fully address the problem of selecting the best relay in a manner that conserves battery power, especially under changing underwater conditions.

The Hop by Hop Vector Based Forwarding (HHVBF) method generates a forwarding vector at each hop instead of relying entirely on a single, fixed source to sink vector[4]. This feature enables local adaptation and can therefore enhance delivery in cases where the distribution of nodes is irregular. Vector based routing still serves as a useful baseline since it involves directional forwarding without the need for complete route construction[2], [4]. However, when several nodes satisfy the forwarding criterion, HHVBF may still result in redundant communications. Furthermore, the method does not specifically protect low energy relays or select forwarders on the basis of expected void risk, forwarding latency, and SNR. In sparse or mobile UWSNs, redundant transmissions waste energy and may reduce the lifetime of nodes that are repeatedly selected. This limitation supports the need for relay selection that evaluates communication quality, battery condition, and future connectivity together.

GEDAR combines geographic forwarding, opportunistic relay selection, and depth adjustment to recover from routing voids [1]. Its contribution is important because it recognizes that packets may reach regions where no suitable next hop is available. Recent UWSN routing studies still treat void handling as a major requirement because void regions increase packet loss, retransmission, delay, and energy consumption [2],[12]. However, recovery after a void occurs is different from estimating void risk before relay selection. A relay may appear suitable at the current hop but may move the packet toward poor onward connectivity. GEDAR does not fully combine anticipated void risk with residual energy, SNR based link quality, forwarding delay, and adaptive reward control. This is the type of coordination required by the proposed DRL formulation.

Opportunistic routing improves reliability by allowing more than one candidate relay to participate in forwarding. The CED-OR mechanism proposed by Ashraf et al. is relevant because it focuses on confined energy depletion and network lifetime in underwater opportunistic routing [11]. It also follows a useful related work pattern by discussing previous protocols and identifying their energy related limitations. However, opportunistic forwarding can create duplicate transmissions or coordination overhead if relay participation is not controlled carefully. More candidates relay could improve the chances of delivery, but if unnecessary or weak relays keep working, they might also cause higher energy use. This means that a stronger system for evaluating relays is needed in opportunistic routing. Instead of depending solely on candidate availability, the proposed DRL framework evaluates each candidate utilizing residual energy, SNR-based connection quality, depth advancement, latency, and void risk.

2.2 Energy Aware and Link Aware Routing Studies

Because early relay node death can split the network, recent energy-aware routing studies prioritize load balancing and residual energy [13], [3], [14]. These results back up the idea that node battery life should not be viewed as a secondary metric. However, energy alone cannot select the optimum relay. A node with

a high residual energy may have poor link quality, slow depth progress, significant delay, or limited forward connectivity. Routing that relies solely on residual energy may preserve a node locally while reducing PDR or increasing latency. This is why the suggested study prioritizes energy conservation but not as the sole routing element. Residual energy is integrated with SNR, depth advancement, forwarding latency, and void risk to increase battery life while maintaining communication performance.

Link-aware routing studies address the instability of underwater acoustic communication by considering SNR or other link quality indicators during relay selection. Recent Q learning based link aware routing for mobile UWSNs shows that communication conditions should influence forwarding decisions because acoustic links are affected by attenuation, noise, mobility, and changing channel states [8]. However, selecting the strongest link repeatedly can drain the same relay nodes and create energy imbalance. A reliable link is valuable only when the relay also supports battery preservation, timely forwarding, and future connectivity. Therefore, the proposed DRL framework uses SNR based link quality as an important component, but it prevents link quality from dominating the decision when residual energy or void risk becomes more critical.

Energy preservation has also been pursued through fuzzy relay ranking, power control, and clustering. Pabani et al. use fuzzy decisions based on hop count, neighbor availability, and distance to the gateway [15], while their FAHP formulation structures relay selection as a multicriteria decision [16]. FLRAF extends fuzzy assessment to link quality, residual energy, and forwarding advancement [17]. APCDBRP balances energy expenditure across nodes [14], and APC-EEDBR adds dual-weight adjustment and adaptive transmission power [18]. RFAODV controls route discovery using location and energy information [19], whereas UWRPL and UWMRPL adapt RPL to underwater topology and mobility [20]. Other approaches distribute forwarding responsibility through adaptive clustering [21], combine PSO routing with data reconstruction requirements [22], or optimize cluster heads and routes through metaheuristic search [23], [24], [25]. These methods show that energy balancing and multicriteria decisions are established capabilities. The proposed DRL formulation instead evaluates the long-term forwarding value of feasible relays and explicitly includes delay and void penalties; its benefit must be assessed through the reported comparisons and ablation experiments.

2.3 Learning Based Routing Protocols

QTAR is a Q learning based topology aware routing protocol for UWSNs, also discussed in the related work of Wang and Ben [7]. It selects next forwarder candidates according to topology information and uses implicit cut vertex recognition to avoid decisions that may weaken connectivity. QTAR is an important recent study because it connects reinforcement learning with topology aware relay selection and reports improvements in energy consumption, latency, and network lifetime. Its limitation in relation to the proposed work is that it is not a DRL formulation centered on battery life improvement through the joint use of residual energy, SNR based link quality, depth advancement, forwarding delay, and anticipated void risk. QTAR shows the value of topology awareness, but the proposed study requires a broader energy centered formulation and component wise ablation to examine how each routing factor affects performance.

ADRP DQL is a recent adaptive distributed routing protocol that applies Deep Q learning in underwater acoustic sensor networks. It considers residual energy, depth, and node degree, and it uses a Deep Q Network to estimate routing action values. ADRP DQL is highly relevant because it confirms that DQL can improve energy efficiency, data delivery, and network lifetime in adaptive underwater routing. Geng and Zhang also use DQN routing to balance residual energy and routing distance [5]. However, the proposed work is not justified merely by using DQL. Its contribution is the energy aware relay formulation that combines residual energy, SNR based link quality, depth advancement, forwarding delay, and anticipated void risk. It also introduces adaptive reward balancing so that routing priorities can change when energy scarcity, weak links, delay, or connectivity risk becomes dominant.

Other learning based routing studies, including reinforcement learning based opportunistic routing and multiagent routing for underwater optical networks, also demonstrate the usefulness of learning for dynamic underwater routing [12], [10]. These methods move beyond static forwarding rules and allow routing behavior to respond to observed network conditions. The optical setting differs from the acoustic network considered here. Some learning based protocols use fixed reward priorities, whereas QLAR already adapts metric weights [8]. The role of routing factors in UWSNs varies throughout time. When batteries are exhausted, energy becomes more critical, link quality becomes more vital, acoustic circumstances deteriorate, and void risk increases as topology gets sparse. A set incentive structure may not adapt well to these changes. This encourages adaptive reward balance in the proposed DRL architecture.

Dynamic DRL and DROR are considered as learning-based baselines in the current evaluation since they show adaptive and reinforcement learning-based routing behavior. DROR integrates depth information, reinforcement learning, and opportunistic forwarding [6]. Their inclusion is critical for determining whether the proposed strategy improves over traditional or opportunistic protocols. However, the crucial point is not whether learning is used, but how the learning policy is evaluated. A DRL policy that does not take into account residual energy, link quality, depth advancement, latency, and void risk may benefit one measure while harming another. Similarly, an opportunistic learning policy can enhance PDR while increasing energy costs if candidate relays are not managed. The suggested strategy differs in that it prioritizes node battery life and network lifetime while preserving dependable delivery and future connectivity.

Recent protocols have linked reliability to connectivity and communication costs. CLOR combines cross-layer relay assessment and coded burst transmission for sparse networks [26], ROBINA and TANVEER handle path disruption with geometric forwarding changes [27], [28]. DIEER coordinates sink mobility, depth thresholds, and holding behavior to ensure delay-sensitive delivery [29]. OCMR predicts current-driven node movement while choosing alternate relays [4], REOVA integrates planned deployment, energy, link quality, and neighbor availability [30]. QLAR forecasts link expiration time and modifies reward weights using the Entropy Weight Method [8], QDER integrates residual energy, depth, and link quality in distributed Q-learning [9]. At the MAC layer, DAWPC-MAC uses DQL for power regulation and collision avoidance [31], GTR integrates trust assessment with Q-learning routing [7]. These qualities preclude adaptive weighing, energy awareness, and void handling from becoming new on their own. The proposed study focuses on their specified integration inside DRL relay selection and evaluates the energy, connection quality, delay, void, and adaptive reward components separately, but does not claim the mobility prediction, security, or MAC techniques provided in those protocols.

Taken together, the reviewed methods show that UWSN routing has advanced from simple progress based forwarding to more adaptive and learning based decision making. However, the existing protocols usually emphasize selected aspects of routing. Depth based and vector based protocols emphasize progress toward the sink, void aware methods emphasize recovery, opportunistic protocols emphasize delivery probability, and learning based protocols emphasize adaptation. The proposed study keeps these concerns connected but gives priority to battery preservation. This is why residual energy, SNR based link quality; forwarding delay, depth advancement, and routing void risk are treated together in one DRL relay selection formulation. Below table, 1 shows the broader comparisons of routing protocols.

The reviewed protocols address different parts of the underwater routing problem rather than the same set of conditions. Conventional schemes mainly rely on depth, hop information, geographic progress, or forwarding vectors, while later approaches introduce void handling, link awareness, and learning based decisions. These developments improve routing in specific situations, but energy preservation is often considered alongside only a limited number of relay factors. This comparison therefore supports the need to assess residual energy together with SNR based link quality, depth advancement, forwarding delay, and void risk when selecting the next relay. The protocol comparison and the response of the proposed study to each limitation are presented in Table 1.

Table 1: Comparison of UWSN routing protocols reviewed in the proposed study

Protocol	**Routing approach**	**Factors**	**Strength**	**Limitation**	**Proposed study addresses the limitation**
DBR	Depth based	Depth	Simple location free forwarding	Does not jointly consider energy, link quality, delay, or void risk	Assesses energy, SNR, delay, and void risk alongside depth before selecting one eligible relay.
H2DAB	Hop based addressing	Dynamic hop address	Supports forwarding without precise localization	Address maintenance adds overhead; relay quality is not jointly assessed	Uses current neighbor states and joint relay assessment; it does not require hop-address maintenance.

Protocol	Routing approach	Factors	Strength	Limitation	Proposed study addresses the limitation
FBR	Focused geographic	Location and transmission power	Limits unnecessary flooding	Power adjustment can increase energy consumption in sparse regions	Filters unsuitable relays before forwarding and penalizes failed attempts to limit wasted transmission energy.
HHVBF	Vector based	Per hop forwarding vector	Adapts forwarding direction locally	May cause redundant transmissions and uneven energy use	Selects one eligible relay using energy and link information to limit redundant forwarding and repeated relay use.
GEDAR	Geographic and opportunistic	Geographic progress and depth adjustment	Provides recovery from routing voids	Primarily reacts to voids rather than predicting relay risk	Includes estimated void risk before forwarding, together with energy, link quality, delay, and depth advancement.
CED OR	Opportunistic	Candidate relay participation and energy considerations	Targets confined energy depletion and network lifetime	Candidate coordination may introduce duplicate transmissions and overhead	Filters and ranks candidates, then selects one relay to limit duplicate forwarding and unnecessary participation.
QTAR	Q learning and topology aware	Topology information and connectivity	Uses learning to support topology aware relay selection	Does not use the full energy centered DRL state considered in the proposed study	Adds SNR, delay, and void-risk assessment to energy and depth in a DRL policy with adaptive reward balancing.
ADRP DQL	Deep Q learning	Residual energy, depth, node degree	Adapts routing decisions using a DQN	Uses a narrower set of relay factors than the present framework	Extends relay assessment with SNR, forwarding delay, and void risk, and adjusts reward priorities as conditions change.
Link aware Q learning	Q learning and link aware	SNR or link quality	Responds to changing acoustic link conditions	Strong links alone may repeatedly burden the same relays	Balances link quality against residual energy, depth, delay, and void risk to avoid repeatedly burdening strong-link relays.

2.4 Research Gap

The reviewed literature shows that each routing category improves a different part of the UWSN routing problem. DBR and H2DAB reduce localization requirements, FBR and HHVBF improve forwarding direction, GEDAR addresses void recovery, opportunistic routing improves delivery probability, and learning based methods improve adaptation. However, the main unresolved issue is how to preserve node battery life while maintaining reliable delivery, acceptable delay, and future forwarding connectivity. Recent methods already combine several routing factors and, in QLAR, adapt their weights [8], [9], [17]. The proposed study therefore examines the specific DRL formulation that coordinates residual energy, SNR

based link quality, depth advancement, forwarding delay, and routing void risk, with component wise ablation used to assess their contribution to battery preservation.

the proposed study addresses this gap by proposing an energy aware DRL relay selection framework for UWSNs. The method evaluates each candidate relay using residual energy, SNR based link quality, depth advancement, forwarding latency, and void risk. It also uses adaptive reward balancing so that routing priorities can change according to network conditions. The comparison with DBR, H2DAB, GEDAR, FBR, HHVBF, ADRP DQL, Dynamic DRL, DROR, and QTAR is therefore necessary to determine whether the proposed formulation provides a better balance between energy preservation, reliable delivery, delay reduction, and future connectivity. The ablation experiments support the evaluation by showing how each part affects the routing behavior.

3. Proposed DRL-Based Adaptive Routing Framework

Underwater wireless sensor networks present unique challenges due to their dynamic environments, limited energy resources, and reliance on acoustic communication, which is often unreliable. To address these issues, this study proposes a novel approach utilizing Deep Reinforcement Learning. This method enables intelligent routing decisions by considering multiple critical factors, including packet delivery ratio, data throughput, node depletion, energy consumption, and network workload. In contrast to existing methods that rely on fixed or pre-determined relay nodes, the proposed approach dynamically selects relay nodes in real time. It adapts to the current network topology, variations in acoustic signal quality, and the residual energy of each node. The fundamental components of the routing problem are outlined, followed by a comprehensive analysis of the adaptive relay-selection model, efficient packet delivery mechanisms, algorithm implementation, operational procedures, baseline protocol selection, performance metrics, simulation parameter settings, and the simulation environment configuration.

The routing framework begins by constructing a detailed representation of the underwater network as a connected graph. This graph model serves as the foundation for all subsequent routing decisions and is formally defined in Equation (1).The network includes a group of sensor nodes that have been placed and possible acoustic communication connections between these nodes.

$$G = (V, E),\ V = \{v_1, v_2, \dots, vN\},\ E = \{(vi, vj) \mid dij \leq Rt\} \qquad (1)$$

Where V is the set of deployed underwater sensor nodes, E is the set of feasible acoustic links, dij is the distance between nodes vi and vj, and Rt is the acoustic transmission range. A sink that is on the surface is put close to where water and air meet, and it gets data packets by passing them through multiple steps. One thing to remember is that the nodes being used can change and might not always be nearby or have enough power. Because of this, a forwarding node can't depend on a single, fixed next hop to send data. It needs to find a new set of relays each time it decides to forward something.

The candidate relay set for any forwarding node is created by carefully checking which neighbors meet the necessary requirements and removing those that don't. Equation (2) explains how this selection works, showing which nearby nodes are allowed to take part in the DRL decision-making process.

$$Ci = \{vj \in V \mid dij \leq Rt, Eresj \geq Emin, LQij \geq LQmin, Progj > 0\} \qquad (2)$$

Here Ci is the candidate relay set of node vi, $Eresj$ is the residual energy of candidate node vj, $LQij$ is the acoustic link quality between vi andvj, and $Progj$ represents positive forwarding progress toward the surface sink. A node is excluded when it has insufficient energy, a bad connection, or has not made any satisfactory progress. When Ci is empty, the packet is regarded as having entered a void area or encountered a routing problem, and the policy receives a penalty before it attempts to look for neighbours once more.

A detailed block diagram together with a topological layout is used to illustrate the practical application of this framework regarding underwater communication, showing how the nodes are positioned at random, how they locate nearby nodes, how possible relay nodes are selected, how the best relay node is chosen to forward the data, the manner in which the data travels from the surface to the sink, and which areas are more likely to have coverage gaps. The layout as well as the method of data transmission between the devices is given in Figure 1.

.

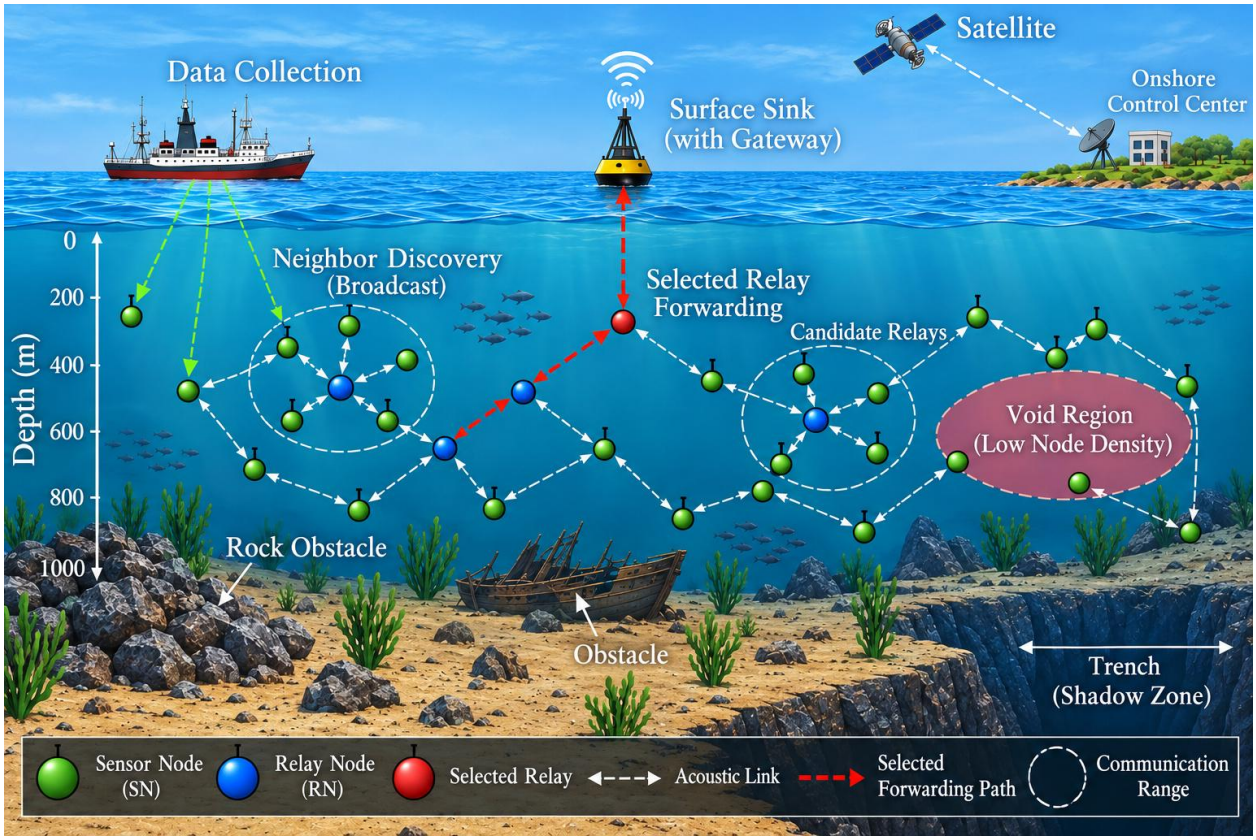


Figure 1. Under Water Sensor Network topology

3.1 Baseline Protocol Selection and Performance Metrics

Algorithm 1 presents the complete procedure, detailing all required inputs, including the underwater sensor network layout, surface sink location, acoustic signal range, initial node energy, learning parameters, number of simulation rounds, and random seeds for each run. The algorithm outputs the learned routing policy, selected relay decisions, comprehensive process records, and complete performance metrics.

Performance evaluation employs multiple metrics to comprehensively assess routing protocol effectiveness. These metrics include network reliability, measured by packet delivery ratio; data transmission speed, assessed through end-to-end latency; energy efficiency, evaluated by residual energy and energy consumption per packet; and the protocol's ability to handle voids, quantified by the number of void failures. Statistical reliability is established using confidence interval calculations across multiple independent simulation runs. The evaluation metrics are summarized in Table 2.

Table 2: Performance metrics

Metrics	Equation	Purpose
Packet delivery ratio	$PDR = (Preceived / Psent) \times 100$	Reliability of successful delivery
End-to-end delay	$Delay = Treceive - Tsend$	Time required to deliver a packet
Residual energy	$Eresidual = Einitial - Econsumed$	Remaining network energy after forwarding
Energy per packet	$EPP = Etotal / Preceived$	Energy cost of successful delivery
Routing overhead	$RO = Pcontrol / Pdata$	Control-packet cost of routing
Throughput	$Throughput = Delivered\ data / Simulation\ time$	Useful delivered data rate
Void failure	$VPF = count(DROP_VOID)$	Number of packets affected by void region
Statistical confidence	$CI95 = 1.96 \times \sigma / \sqrt{K}$	Multi-seed result reliability

3.2 Experimental Setup

The same simulation settings are used in all the tests so that both the proposed DRL framework and each of the baseline methods can be evaluated under identical conditions with respect to deployment, traffic,

communication channels, and energy consumption. The NS-3 simulation platform, with its capability of supporting underwater acoustic routing, is used as the foundation for setting up three-dimensional underwater wireless sensor networks in which the nodes can be positioned anywhere. The surface sink is located at or near the point where the water meets the air and receives the data packets through a number of sound-based forwarding steps. Network density is examined in various scenarios involving different numbers of nodes in order to assess how well the system scales and performs consistently. A summary of the simulation settings is given in Table 3.

Table 3: Experimental Structure Setup.

Parameters	**Setting**
Simulation platform	NS-3 simulator configured with underwater acoustic routing capabilities.
Network type	Three-dimensional topology for the underwater wireless sensor network
Node deployment	Sensor nodes are placed arbitrarily within the underwater environment
Sink position	A surface sink is positioned at or near the water-air interface
Primary node density	Network configuration includes 100 sensor nodes
Scalability densities	Experiments are conducted with 50, 100, 150, and 200 sensor node
Communication mode	Multi-hop acoustic forwarding is employed for data transmission
Traffic model	Source nodes generate packets periodically
Energy model	Energy consumption is measured for transmission, reception, forwarding, and processing operations
Environmental factors	Environmental factors considered include temperature, salinity, pressure, noise, turbulence, and node mobility
Independent runs	Simulations are repeated using multiple random seeds K
Compared protocols	DBR, H2DAB, GEDAR, FBR, HHVBF, ADRP-DQL, Dynamic Clustering DRL, DROR, QTAR, and Proposed DRL

The simulation environment enables packet forwarding to take place in a consistent manner under the same underwater setup and acoustic channel conditions for each of the protocols. The source nodes continue to send packets, identify which of the nearby nodes are available, determine which ones could be used to forward the packets, select the best ones to send the data to, and attempt to get the packets to the main sink node at the top. The simulator logs events at the packet level, including the sending of packets, moving them on their way, receiving them, exchanging control messages, experiencing link failures, handling void-region problems, and keeping track of the amount of energy consumed. These trace records are then examined in detail in order to establish all the performance metrics employed to evaluate the system. The events recorded during the simulation serve as the connection between the original data from the simulation and the actual performance measurements. The events relating to packet creation and packet sending are used to determine the total load and the degree to which packets are delivered. The intermediate nodes are monitored when they forward data, which aids in understanding how the paths function and how efficiently the data is being sent. The surface sink reception events are used to count the number of packets that are delivered successfully, which in turn allows the delay and throughput to be calculated. The routing control messages are tracked in order to assess the amount of overhead they introduce. Void-region failures occur when there are no possible relay options available, illustrating how well voids are detected and managed. When acoustic links fail after forwarding attempts, this shows how reliable the links are and how well the system recovers from failures. The records of node energy consumption are used to analyse how much energy remains and how much energy is used for each packet that is successfully delivered. A summary of the recorded trace events and their roles with respect to the metrics is given in Table 4.

Table 4: Packet trace events and their metric roles.

Trace event	**Meaning**	**Metric role**
SEND	Packet created and communicated by a source node	PDR denominator and offered load
FORWARD	Intermediate node forwards a packet	Forwarding count and path behavior
RECEIVE	Surface sink receives the packet	PDR numerator, delay, throughput
CONTROL	Routing control message exchanged	Routing overhead
DROP_VOID	No feasible candidate relay exists	Void-region failure and penalty
DROP_LINK	Acoustic link failure after forwarding	Link reliability and failure penalty
ENERGY	Node energy consumption recorded	Residual energy and energy per packet

The method presented is tested under various network conditions in order to examine how stable and flexible it is when dealing with different underwater environments. In normal deployment situations, with typical node spacing and a consistent sound environment, the system's signal routing is set up in the standard way. The system is put to the test in situations where there are only a few relays available by considering the risks involved when no relays are available. It is important to check how effective the energy-aware relay filtering is when some of the nodes run out of energy. Testing is also carried out in random configurations where the nodes are placed in irregular patterns to see how well the system can cope with any kind of network layout. High mobility scenarios with node movement due to currents and turbulence examine robustness under changing topology conditions. The scenario conditions and their purposes are summarized in Table 5.

Table 5: Scenario working used for methodology validation.

Scenario	Condition	Purpose
Normal deployment	Standard node density and stable acoustic conditions	Basic routing behavior
Sparse deployment	Nodes are far apart and candidate relays are limited	Void-risk handling
Dead-node condition	Some nodes lose energy or become unavailable	Energy-aware relay filtering
Random deployment	Nodes are irregularly distributed in the field	Adaptability to arbitrary topology
High mobility	Nodes move due to currents or turbulence	Robustness under changing topology

Performance values are averaged over independent random seeds. The 95% confidence interval for each metric is calculated underneath given equation, thereby ensuring that the reported performance is not derived from a single successful run.

$$CI95 = 1.96 \times \sigma / \sqrt{K} \quad (3)$$

The σ signifies the standard deviation of a metric across K independent runs. The same statistical method is applied to both the proposed DRL framework and each baseline protocol. This methodology is completed by associating the proposed adaptive relay-selection process, the settings of simulation parameters, and the collection of traces, the operational scenarios, and the computation of metrics within a single reproducible workflow.

3.3 Adaptive Relay Selection Strategy

The formation of the relay set through the filtering process, the current node proceeds to observe the routing condition of each candidate relay. This observation is captured in a comprehensive state vector that serves as input to the DRL decision module.

$$Sj = [Eresj, LQij, DPj, DistPj, Dij, VRj, FLj, ERj] \quad (4)$$

The state vector combines multiple critical parameters including residual energy, acoustic link reliability, and topological progress toward the sink, expected delay characteristics, void-risk assessment, load-balancing considerations, and environmental reliability information. The detailed state representation enables the DRL agent to make wise decisions regarding which relay to use when taking into account various goals at the same time. The action space, as given in Equation (5), consists of all the possible options for selecting a relay from those available.

$$Ai = \{aj \mid aj = select(vj), vj \in Ci\} \quad (5)$$

Each action corresponds to selecting one specific candidate relay node for forwarding the current packet. A key aspect of the proposed method is that the selected relay is not predetermined; rather, it is dynamically chosen as the optimal option for the current forwarding scenario. This approach demonstrates the flexibility and responsiveness of the method.

3.4 Packet Delivery Optimization Process

Enhancing packet delivery extends beyond the design of reward mechanisms and requires a comprehensive system for monitoring and evaluating performance. One thing to note is that Equation (6) shows the packet-delivery score, which is used to evaluate each candidate relay node vj.This score takes into account both good and bad factors.

$$Rj = 0.25Eresj + 0.20LQij + 0.15ERj + 0.10DPj + 0.10DistPj - 0.10Dij - 0.07VRj - 0.03FLj \quad (6)$$

Candidates who have more leftover energy, a stronger connection in sound, are more reliable in their environment, show more confidence in how deep they go, and move forward more positively in distance get better scores. On the other hand, the scoring system gives lower points to nodes that have long delays in forwarding messages, a high chance of being a dead end, or are used too often as relay points. If packet forwarding doesn't work because of a broken acoustic link or routing problems, a penalty is applied, and the process of finding neighboring nodes is done again, which helps achieve the goal of delivering packets in the routing system. The last choice for the relay is decided by the value from the DRL model, as shown in Equation (7), which picks the candidate that gives the highest Q-value.

$$v* = arg\ maxvj \in Ci\ Q(Sj, aj;\ \theta) \qquad (7)$$

Where $v*$ is the selected relay node, $Q(Sj, aj;\ \theta)$ is the estimated policy value for selecting candidate vj under state Sj, and θ denotes the learnable model parameters. Once the packet is forwarded, the policy is adjusted based on the reward or penalty given for delivery.

$$Q(s,a) \leftarrow Q(s,a) + \alpha\,[r + \gamma\, maxa'\, Q(s',a') - Q(s,a)] \qquad (8)$$

The learning process uses α as the learning rate, γ as the discount factor, r as the immediate reward after taking an action, s as the current situation, a as the chosen action to forward the signal, and s' as the new situation that results after the action is taken. This equation allows the routing policy to learn from both successful packet delivery and failed forwarding events.

A comprehensive algorithm has been developed to formalize the complete working procedure of the proposed DRL model. The algorithm commences when an underwater source node produces a data packet destined for the surface sink. Prior to relay selection, the model compiles fundamental network information including node positions, residual energy levels, neighboring relationships, link conditions, and mobility-related data. The current node identifies neighboring nodes within acoustic transmission range as potential forwarding options, though not all neighbors qualify as suitable relays.

The commencement of the process occurs when an underwater source node produces a data packet. This source node can be any node positioned within the underwater domain. Before a relay is chosen, the model compiles fundamental network data. This includes the position of nodes, remaining energy levels, neighboring links, link conditions, and information related to mobility. The current node looks for other nodes that are close enough to be reached through its sound signal. These nearby nodes can be used as possible paths to send the data. It's important to remember that not all neighbors can be used as relay nodes. A node needs to be suitable for handling forwarding tasks. The model forms a relay group by verifying whether the adjacent nodes have sufficient energy remaining and whether they are able to forward the data closer to the sink; if the relay group is empty, then the current node cannot forward the packet any further and this indicates that the node is located in a void area or that no suitable forwarder is available.

If there are candidate relays available, the algorithm then proceeds to observe and evaluate the current state. The scoring system awards more points to relays that are reliable, have sufficient energy, and are able to transmit data smoothly to the sink; it assigns fewer points to relays that cause delays, are risky since they may not work well, or are overloaded. This reflects the primary objective that the proposed method is aiming to achieve. The algorithm then calculates the rewards for all the candidate relays and selects the one with the best routing value. After a relay has been chosen, it is subjected to a check to ensure that it is still accessible, functioning properly, and not directed in a way that would prevent signals from reaching their destination.

It is important to carry out this step since the underwater environment has the potential to change rapidly due to moving objects, unstable connections, or low energy levels. When the relay that has been selected fails, it is removed from the list of possible relays and the algorithm looks for a different relay to use. After the packet has been confirmed as valid, it is forwarded to the chosen relay node and the process goes on like this until the packet finally reaches the surface sink.

Each time a packet is forwarded, the algorithm checks whether the packet has reached the sink or has been sent to the next relay; good feedback is obtained when a delivery is made. But if it fails, there is a penalty, and the failed relay option is removed from the list of candidates. This system helps the model stop picking relays that don't work well over time.

Throughout the forwarding process, the simulation logs various trace events including packet transmission, forwarding operations, reception acknowledgments, void-related drops, link failures, These trace events

serve as the foundation for calculating performance metrics such as packet delivery ratio, end-to-end delay, residual energy distribution, energy consumption per packet, routing overhead, throughput, void packet failures, and network lifetime and energy consumption records.

Algorithm. Proposed DRL-Based Adaptive Relay Selection and Packet Forwarding

Input: UWSN topology G, surface sink, acoustic range Rt, initial node energy, learning parameters α and γ, simulation rounds T, independent seeds K.

Output: Learned routing policy, selected relay decisions, trace records, and performance metrics.

Initialise arbitrary underwater nodes, surface sink, node energy, routing tables, and learning policy.
for each independent seed k = 1 to K do
  for each simulation round t = 1 to T do
    for each generated packet p at current node vi do
      Discover neighbours and construct candidate relay set Ci using Eq. (2).
      if Ci is empty then
        Declare void region / routing failure; assign negative penalty Rvoid.
        Update policy using Eq. (7) and repeat neighbour discovery for packet p.
      else
        for each candidate relay node vj in Ci do
          Observe state vector Sj using Eq. (3).
          Compute packet-delivery feedback score Rj using Eq. (5).
        end for
        Select action aj from the action space in Eq. (4).
        Select relay node v* using Eq. (6).
        Forward packet p to selected relay node v*.
        if packet p is delivered successfully then
          Record RECEIVE event; assign positive delivery reward.
          Update routing policy using Eq. (7); stop forwarding for packet p.
        else
          Record DROP_LINK or DROP_VOID event; assign failure penalty.
          Update routing policy using Eq. (7) and return to neighbour discovery.
        end if
        Record energy consumption, forwarding load, and trace events.
      end if
    end for
  end for
end for
Calculate PDR, delay, energy, routing overhead, throughput, void failure, and confidence interval.

The flowchart shows the void region branch, the packet delivery failure branch, the penalty assignment, the policy update, and the continuing neighbor discovery process. The process ends only when the packet is

delivered successfully for the current forwarding step. The packet forwarding process is shown in Figure 2.

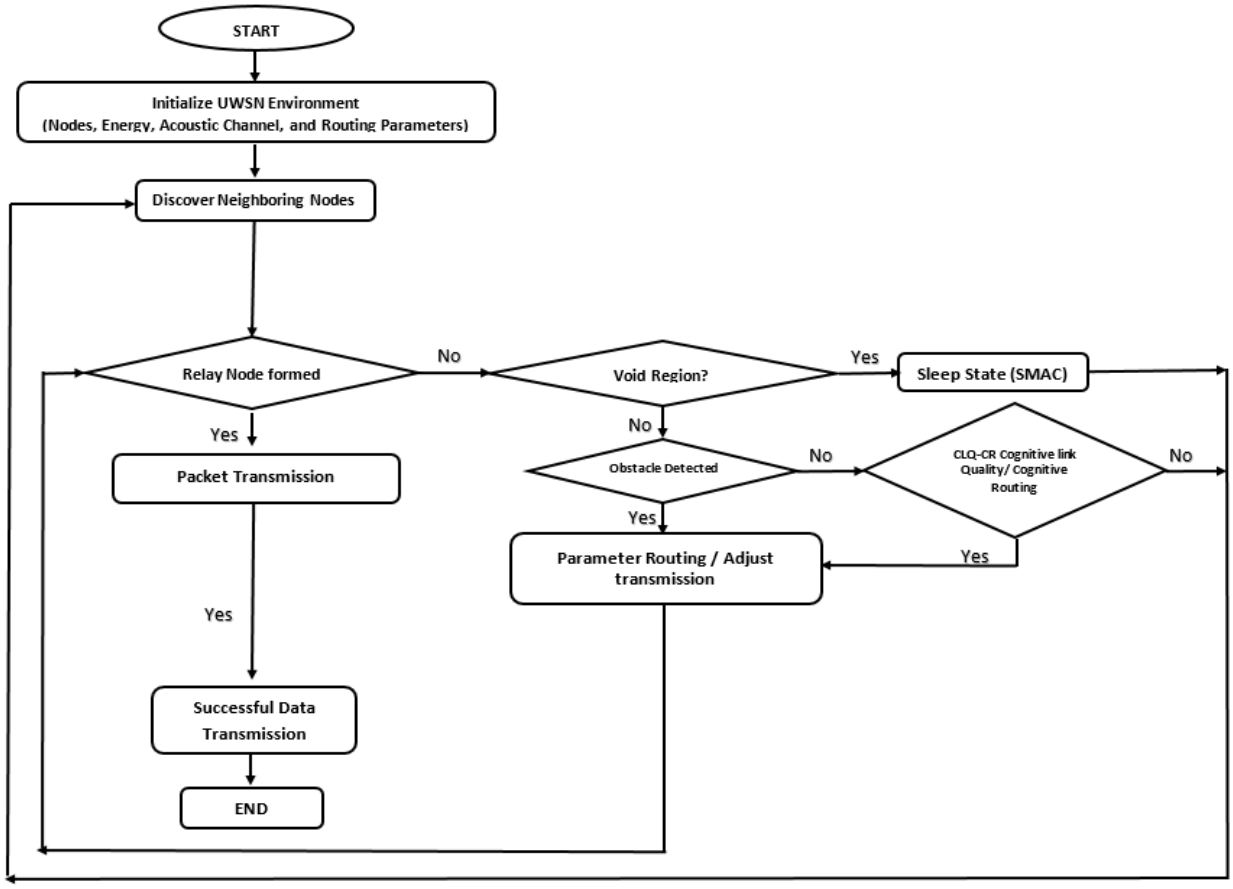


Figure 2. Adaptive relay selection

3.4 Baseline Protocol Selection and Performance Metrics

The full algorithm is shown in Algorithm 1, which lists all the inputs needed, such as the layout of the underwater sensor network, where the surface sink is located, how far the acoustic signals can travel, the starting energy of each node, the learning settings, the number of simulation rounds, and the random seeds used for each run. The algorithm provides the routing policy that it has learned, the relay decisions it made, complete records of the process, and all the performance measurements.

Performance evaluation uses several measures that together check how well a routing protocol works. These metrics include how reliable the network is by looking at the packet delivery ratio, how fast data moves by checking end-to-end latency, how energy-efficient it is by measuring residual energy and energy used per packet, and how well it deals with voids by counting void failure occurrences. Statistical reliability is made sure by using confidence interval calculations in many separate simulation runs. The evaluation metrics are summarized in Table 6.

Table 6: Performance metrics.

Metric	Equation / computation	Purpose
Packet delivery ratio	$PDR = (Preceived / Psent) \times 100$	Reliability of packet delivery
End-to-end delay	$Delay = Treceive - Tsend$	Packet delivery latency
Residual energy	$Eresidual = Einitial - Econsumed$	Residual network energy post-forwarding
Energy per packet	$EPP = Etotal / Preceived$	Energy consumption per successful delivery
Routing overhead	$RO = Pcontrol / Pdata$	Control packet overhead in routing
Throughput	$Throughput = Delivered\ data / Simulation\ time$	Effective data delivery rate
Void failure	$VPF = count(DROP_VOID)$	Packets affected by void regions

Metric	Equation / computation	Purpose
Statistical confidence	$CI95 = 1.96 \times \sigma / \sqrt{K}$	Reliability across multiple seed runs

3.5 Signal-to-Noise Ratio

An essential aspect of underwater sound communication, one that has a major impact on the effectiveness with which messages are transmitted, is the ratio between the power of the signal and the level of background noise. The channels used for underwater sound communication suffer from heavy signal attenuation, are affected by background noise arising from a variety of sources, and have varying path delays for sound waves. The signal-to-noise ratio influences the quality of the acoustic link between nodes and thus has an effect on the reliability of data packet delivery. A higher signal-to-noise ratio means that the signal is clearer and there are fewer errors, resulting in better communication; whereas when the signal-to-noise ratio is low the signal becomes more distorted and this causes a greater number of packet errors with the need for more retransmissions.

The receiving end's signal-to-noise ratio is influenced by a number of factors, such as the strength of the signal when it is transmitted, the extent to which the signal diminishes as it travels over a distance, the level of background noise present, for example thermal noise, noise caused by the movement of water, noise from ships, and noise resulting from the wind on the surface, as well as the frequency of the signal being used. The distance at which sound can be clearly heard is determined by how well the signal is distinguished from the background noise. The greatest distance at which you can reliably communicate is set by the level of noise that makes the signal difficult to understand. If the signal-to-noise ratio falls below the acceptable level, the distance over which communication functions properly decreases, and this may result in network splits or in certain areas where communication is not possible.

The effect of the signal-to-noise ratio in the proposed framework is indicated indirectly by the acoustic link quality parameter given in Equation (2). This parameter reflects how well the signal is distinguished from the noise, considering both the strength of the signal and the kind of noise present in the water; the higher the signal-to-noise ratio, the better the link quality, and therefore such links are more desirable when selecting a relay. Conversely, links having a low signal-to-noise ratio, which indicates poor quality, are eliminated when the relay candidates are being chosen. This prevents unreliable forwarding decisions that could lead to packet loss and the wasting of energy.

The Signal-to-Noise Ratio changes are also considered in the environmental reliability component of the state vector as given in Equation (3), since this shows how stable and consistent the acoustic channel is. Nodes which have a good Signal-to-Noise Ratio and stable channels receive higher scores in the packet-delivery feedback part of Equation (5). By integrating this information, the DRL agent is able to select relays that have a better Signal-to-Noise Ratio, thus improving packet delivery reliability and preventing the agent from wasting too much energy on poor acoustic links.

In order to properly compare the various routing protocols, the different methods had to be tested using the same model of sound propagation, under similar setup conditions, with identical assumptions regarding traffic, and in the same simulation environment. The experiments were carried out using NS-3 and Aqua-Sim, and in them the proposed DRL framework was compared with nine existing routing methods. These methods consisted of traditional protocols such as DBR, H2DAB, GEDAR, FBR and HHVBF, as well as learning-based approaches including ADRP-DQL, Dynamic DRL, DROR and QTAR.This common evaluation setting provides a consistent basis for examining whether multi-factor state assessment and adaptive reward design improve reliability, latency, energy utilization, routing efficiency, and robustness in underwater wireless sensor networks.

Packet delivery performance provides the first indication of the reliability achieved by the proposed routing strategy. The proposed DRL approach attains the highest PDR at 93.00%, compared with 86.00% for ADRP-DQL and 84.00% for Dynamic DRL, while the conventional protocols remain below the learning-based methods. Relative to ADRP-DQL, the proposed approach improves PDR by approximately 8.14%, indicating that a larger proportion of generated packets successfully reaches the destination. This improvement is consistent with the use of signal-to-noise ratio, residual energy, depth advancement, and void risk in relay selection, which helps the policy avoid weak or unsuitable forwarding paths. The comparative reliability advantage of the proposed DRL approach is presented in Figure 3.

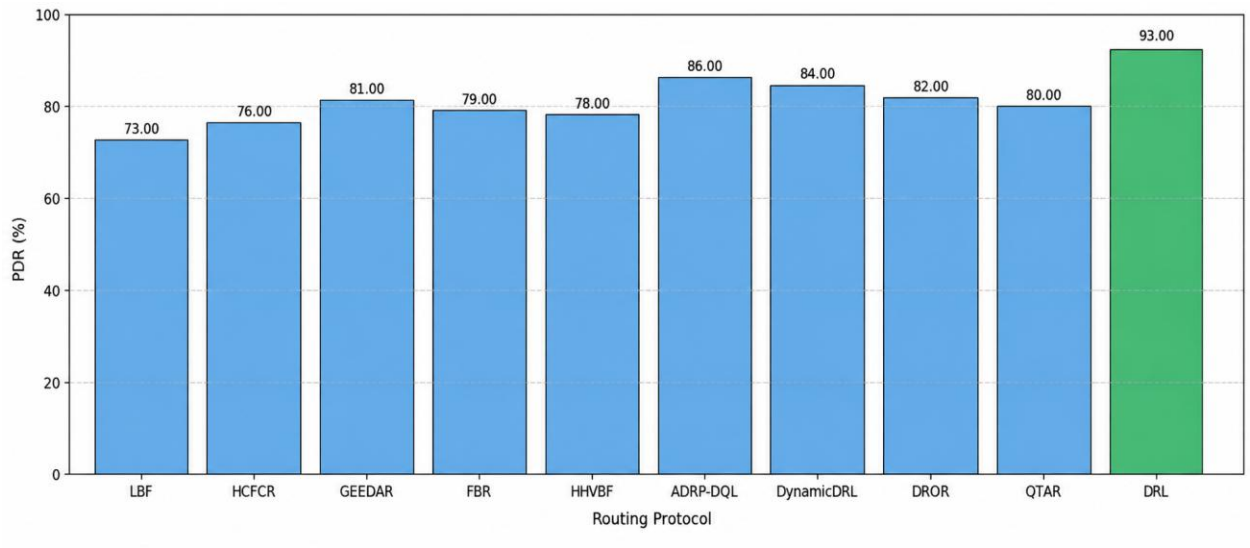


Figure 3. Comparative PDR of the DRL Vs other Routing Protocols.

3.6 Environmental Factor Analysis

Environmental conditions are important in UWSNs because acoustic propagation changes with the physical state of the water. Variations in temperature, salinity, pressure, currents, turbulence, ambient noise, surface conditions, thermoclines, suspended sediment, and other marine effects can alter sound speed, attenuation, node position, signal clarity, and therefore the stability of communication links. A routing method that has been tested only under ideal and unchanging channel conditions could create a misleading impression regarding its reliability in actual use. The purpose of the environmental analysis in this study is to examine how these factors can influence the selection of routes and to explain why environmental reliability should be considered together with energy, delay, link quality, and the risk of voids. The environmental factors that were examined and the way in which they are expected to affect communication are given in Table 7.

Table 7: Effects of Environment on underwater acoustic routing.

Environmental factor	**Behavior**
Water Temperature	Changes sound speed and acoustic signal behavior.
Salinity	Affects sound propagation and link quality.
Water Pressure and Depth	Higher depth increases pressure and affects deployment conditions.
Ocean Currents	Move nodes and create topology changes.
Tides	Change water level and node positions over time.
Turbulence	Creates unstable acoustic links.
Ambient Noise	Reduces signal clarity and increases packet error probability.
Seabed Composition	Affects sound reflection and absorption.
Surface Reflection	Creates multipath signal distortion.
Thermocline Layers	Bends acoustic signals due to temperature layers.
Air Bubbles	Scatter acoustic waves and reduce communication quality.
Marine Life Activity	Adds noise and movement in underwater environment.
Weather Conditions	Affect surface waves and acoustic noise.
Ice Cover	Changes surface reflection and acoustic propagation.
Sediment Suspension	Weakens acoustic communication and signal clarity.

Table 8: Impact scores for water temperature and salinity.

Environment factor	**Impact score (%)**
Water Temperature	78
Salinity	72

The environmental assessment provides an overview of fifteen environmental factors and the way in which they are likely to affect underwater sound communication. The quantitative assessment is more restricted in scope and considers only temperature and salinity as examples of hydrographic variables; the remaining thirteen factors are included in the qualitative section and do not have individual impact scores given in that

table. The chapter fails to explain how these two variables were selected or how their scores were calculated, which in turn makes the quantitative comparison more difficult to understand. These scores should therefore not be interpreted as a ranking of all fifteen factors or as direct percentage losses in PDR or energy. The qualitative effects and representative temperature and salinity scores are reported in Tables 7 and 8, respectively.

Water temperature has a representative impact score of 78%, while salinity records 72%, indicating a stronger reported influence for temperature within this two-variable comparison. Incorporating environmental reliability into the DRL state and reward formulation allows the routing policy to consider changes in acoustic-link reliability when assessing candidate relays. The comparative impact scores are presented in Figure 4.

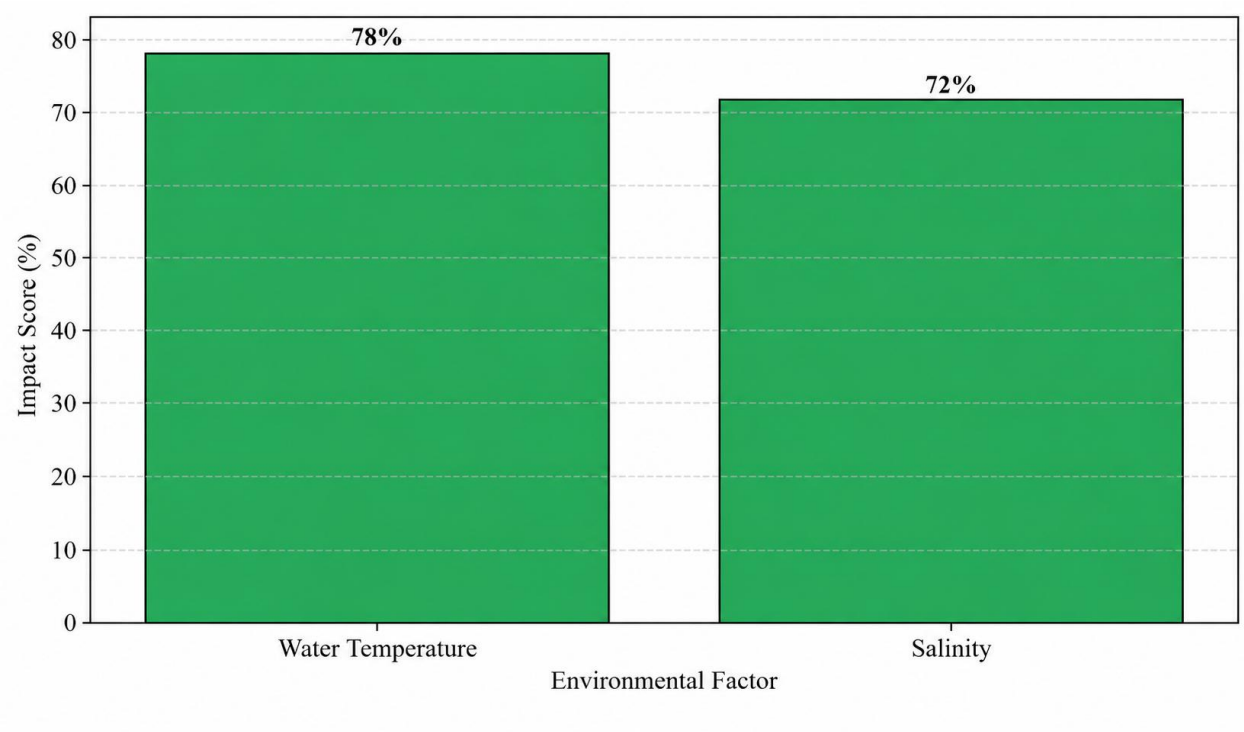


Figure 4. Impact of water temperature and salinity

Robustness across a wider set of environmental disturbances was examined to determine whether the routing advantage persists beyond temperature and salinity alone. The proposed DRL trajectory remains above the competing protocols across the evaluated environmental factors, whereas the baseline methods show greater deterioration as acoustic stress increases. The comparison evaluates the capacity of each protocol to preserve routing robustness when the communication medium becomes less favorable, as shown in Figure 5.

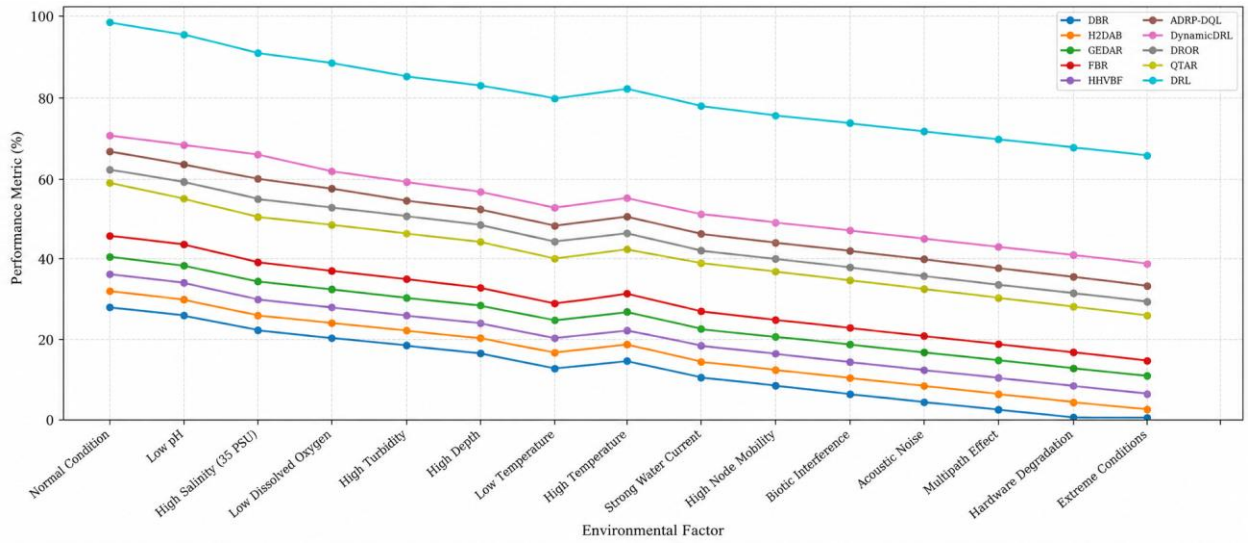


Figure 5. Routing robustness of the evaluated protocols under multiple underwater environmental disturbances.

The packet-delivery analysis provides a direct network-level measure of performance under environmental variation. Across the evaluated factors, the proposed DRL framework maintains the highest PDR trajectory, while the conventional and alternative learning-based protocols experience larger reductions as environmental conditions become more demanding. The resulting PDR behavior is presented in Figure 6.

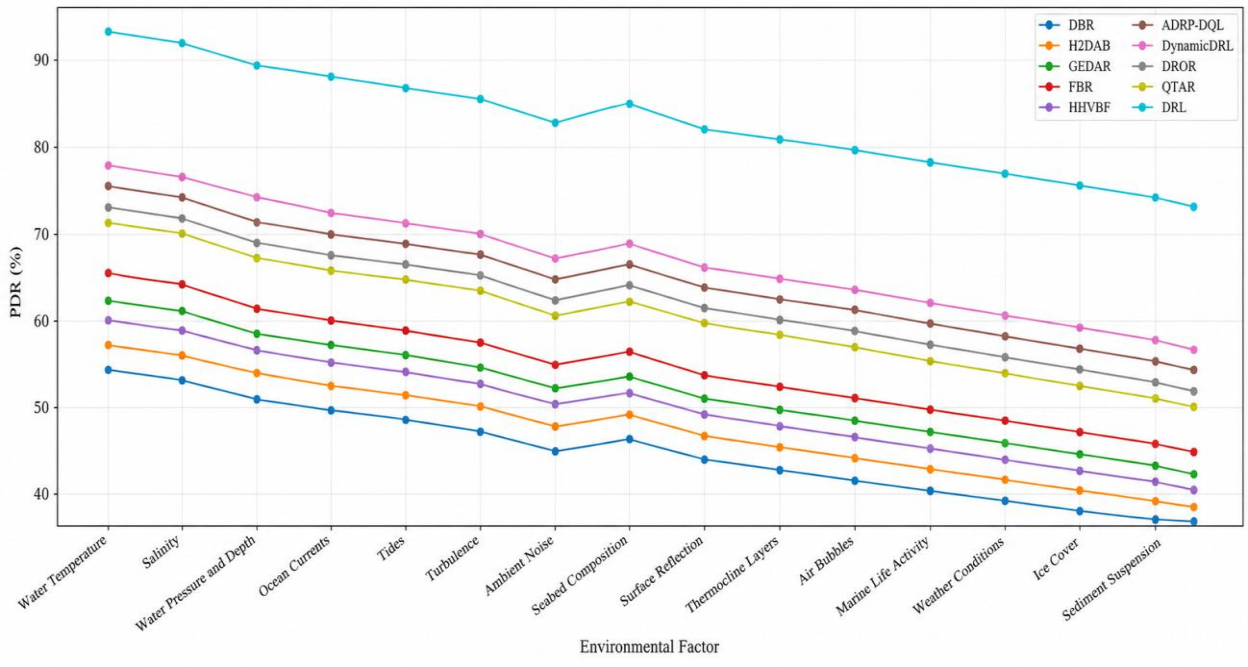


Figure 6. PDR evaluated protocols in changing underwater environmental factors.

The latency results complement the reliability findings by showing that improved delivery is not obtained at the cost of longer packet transfer time. The DRL approach records the lowest end-to-end delay at 0.98 s, whereas ADRP-DQL records 1.16 s and the conventional DBR and H2DAB protocols require 1.65 s and 1.52 s, respectively. Compared with ADRP-DQL, this corresponds to an approximate 15.52% reduction in delay. The lower latency is consistent with the policy's preference for forwarding choices that limit holding time, unreliable hops, void encounters, and subsequent retransmissions. The result therefore suggests that multi-factor relay selection improves both delivery reliability and transmission responsiveness under the evaluated conditions, as shown in Figure 7.

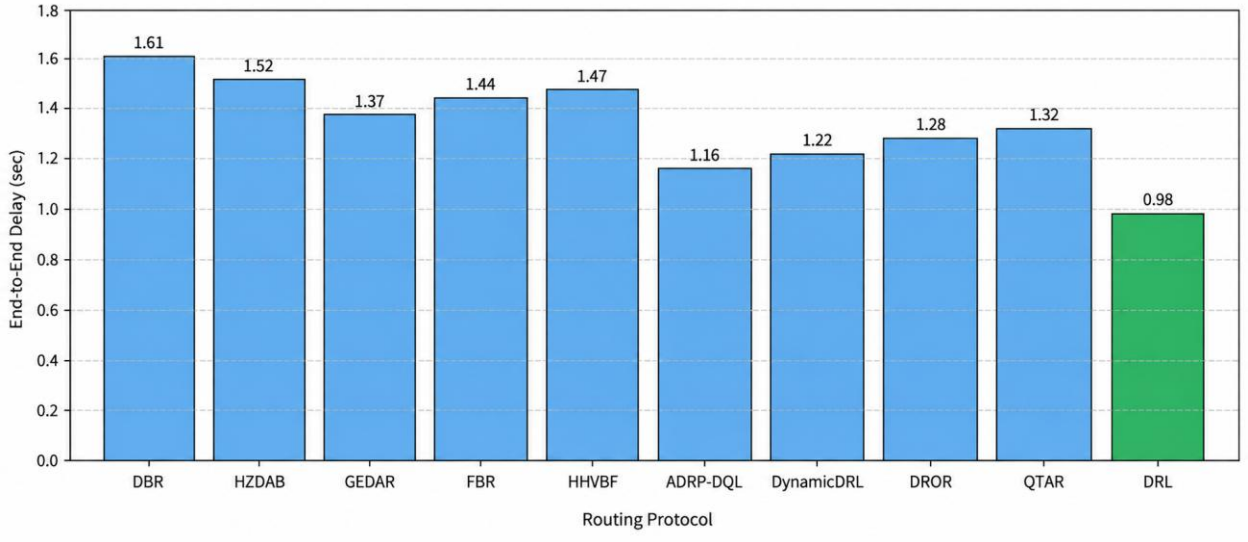


Figure 7. Comparison of end-to-end delay with proposed DRL and other routing protocols.

Energy preservation shows a similar advantage for the proposed routing strategy. The proposed DRL framework retains 87.70% residual energy, exceeding Dynamic DRL at 81.00%, ADRP-DQL at 79.00%, and the conventional DBR protocol at 68.00%. Relative to Dynamic DRL, the strongest competing result for this metric, the proposed method increases retained energy by approximately 8.27%. The fact shown by this pattern is that improved packet delivery results from energy being used more evenly rather than from

trying hard to send data. By taking account of the residual energy, relays which have no power left are avoided and the task of forwarding data is distributed among the nodes which still have sufficient battery power, thus helping to prevent some nodes from becoming overloaded too early. The comparison of the performance with respect to residual energy is given in Figure 8.

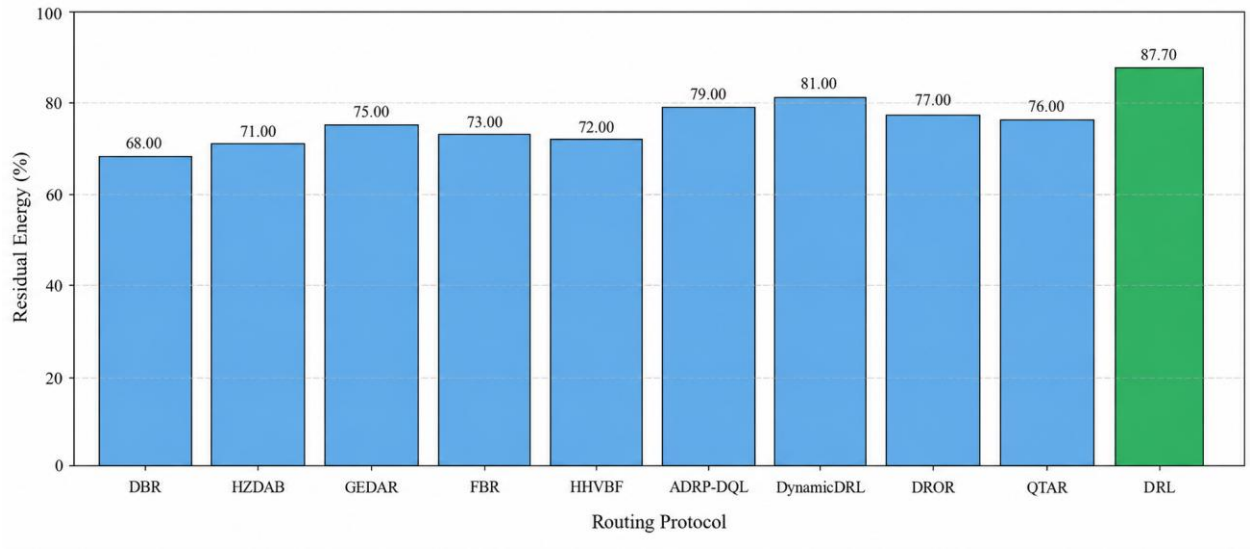


Figure 8. Residual energy (%) preservation achieved by the proposed DRL and other routing protocols.

The drop in energy used for each packet that gets successfully delivered shows that the new routing policy works well. The new DRL model uses 0.41 joules per packet, which is better than ADRP-DQL at 0.60 joules and Dynamic DRL at 0.62 joules.DBR uses 0.86 joules and H2DAB uses 0.80 joules. Compared to ADRP-DQL, the new method uses about 31.67% less energy per packet. This saving happens because weak links and forwarding paths that are likely to fail are filtered out before sending data, which reduces the need for repeated attempts and failed forwards. The measure illustrates the efficiency of communication during the simulation and should be examined by itself, not in relation to any previous learning or offline costs. The comparison of energy per packet is given in Figure 9.

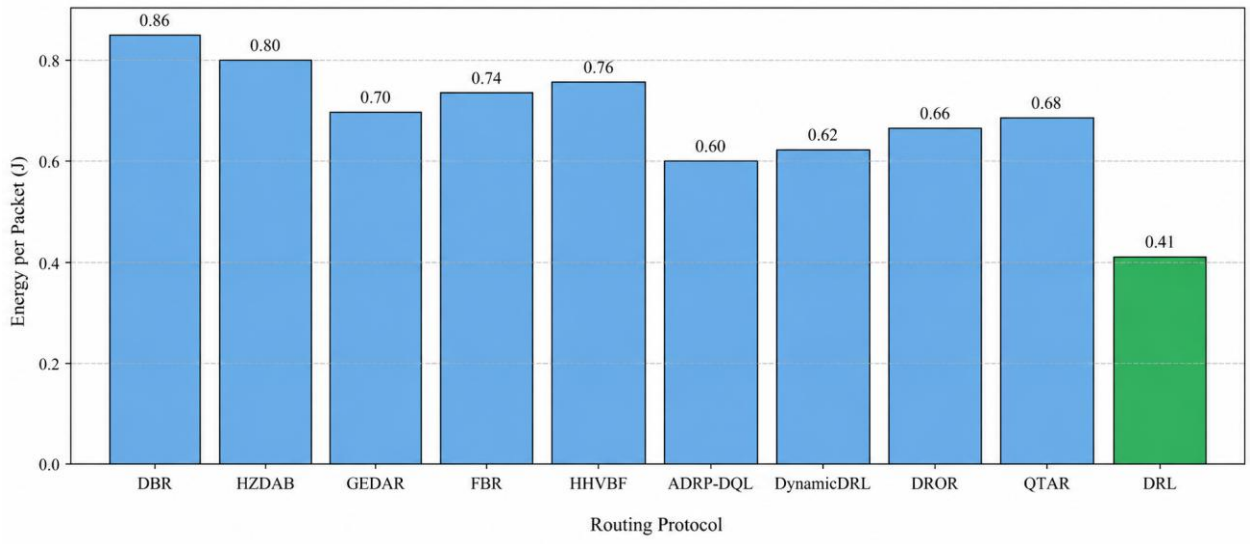


Figure 9. Energy consumed per successfully delivered packet across the evaluated routing protocols.

The longer period that the network can operate is due to the energy savings at the packet level, and the new DRL framework proves to be the most durable, lasting for 662.60 rounds whereas Dynamic DRL lasts for 625 rounds, ADRP-DQL for 610 rounds, and DBR for 510 rounds. This represents an improvement of approximately 6.02% over Dynamic DRL, the closest competing protocol for network lifetime. The result is consistent with the joint consideration of residual energy and forwarding load, which reduces repeated use of critical relays and delays the formation of energy-depleted regions. Network lifetime nevertheless represents an aggregate measure and should be interpreted together with residual-energy trends because localized node exhaustion can affect coverage before complete network termination. The comparative lifetime result is presented in Figure 10.

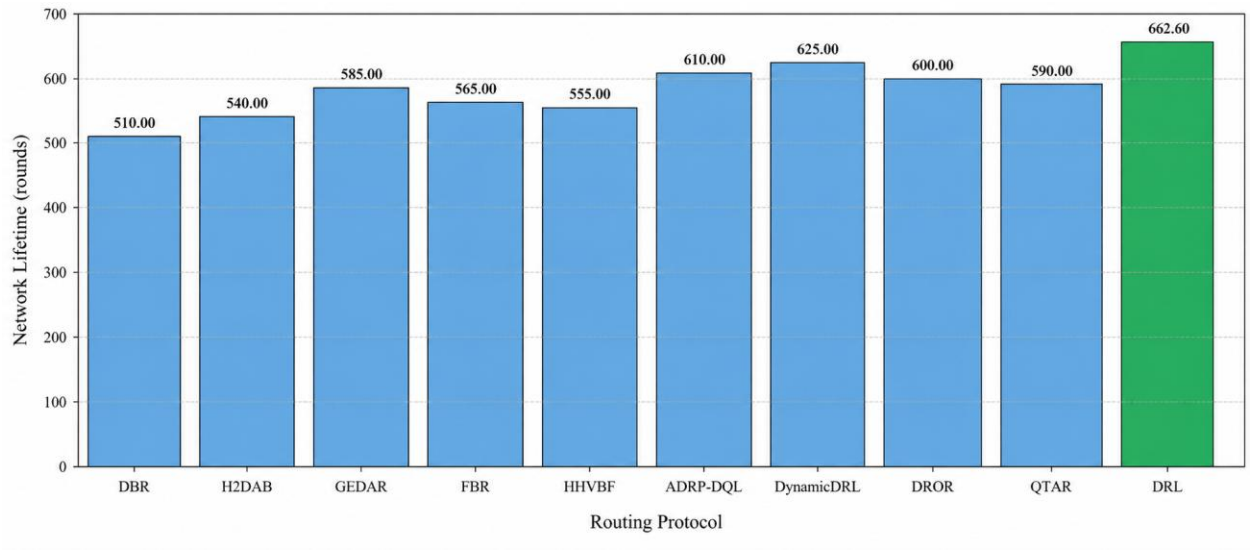


Figure 10. Network lifetime comparison of the proposed DRL and other routing protocols.

Higher delivery reliability and better energy utilization are accompanied by improved network throughput. The proposed DRL framework achieves the highest normalized throughput of 86.00%, compared with 74.00% for ADRP-DQL and 72.00% for Dynamic DRL. Relative to ADRP-DQL, this corresponds to an improvement of approximately 16.22%. The increase is consistent with fewer packet losses, fewer void-induced failures, and more reliable relay selection, which collectively allow a greater proportion of generated traffic to reach the sink. Because throughput is reported here as a normalized percentage, the result describes relative transfer efficiency rather than absolute acoustic-channel capacity. The comparative throughput performance is illustrated in Figure 11.

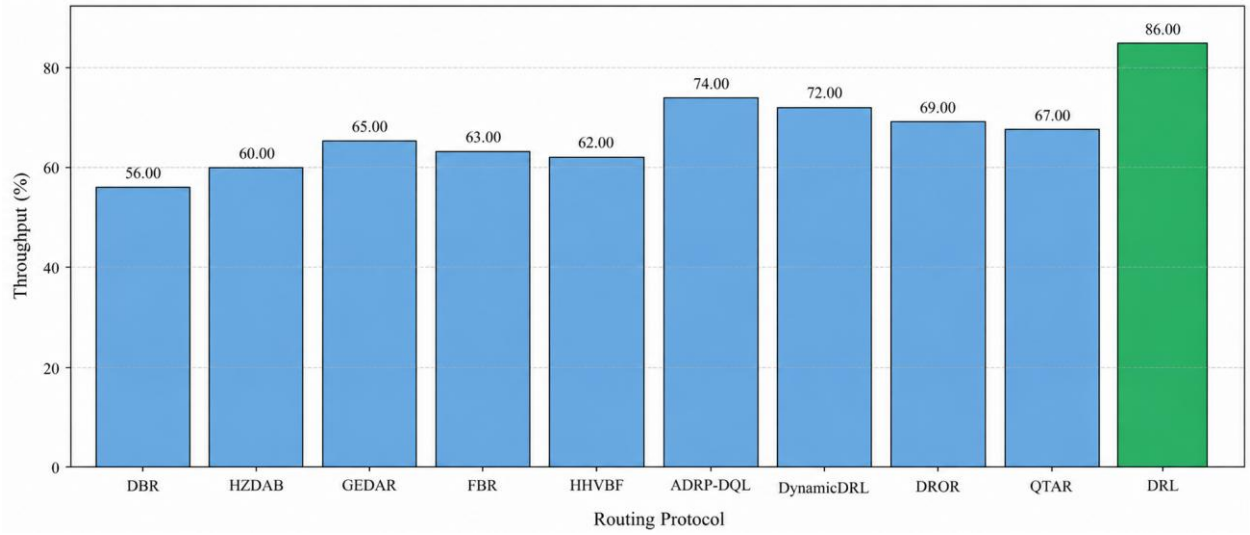


Figure 11. Normalized network throughput achieved by the evaluated routing protocols.

Routing efficiency is also reflected in the amount of control traffic required to sustain packet forwarding. The proposed DRL framework records the lowest routing overhead at 8.50%, compared with 12.40% for ADRP-DQL, 12.80% for Dynamic DRL, and 19.50% for DBR. Relative to ADRP-DQL, this represents an overhead reduction of approximately 31.45%. The lower control burden means that relay decisions are more reliable, which helps reduce route recovery, extra transmissions, and signaling caused by failed forwarding attempts. This network-level gain needs to be thought about together with the extra local computation needed to calculate the DRL state vector, since communication costs and processing costs are different kinds of expenses. The comparison of routing overhead is shown in Figure 12.

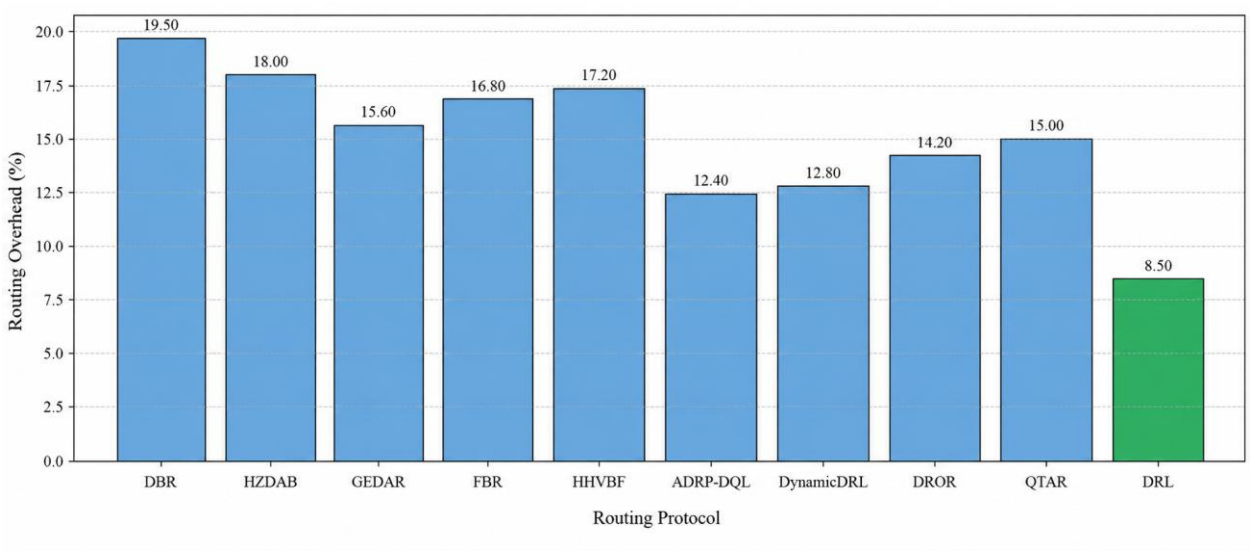


Figure 12. Routing-control overhead of the proposed DRL and benchmark routing protocols.

Resistance to routing voids offers a clearer way to check how stable forwarding is in sparse or irregular underwater network setups. The proposed DRL framework has the lowest rate of packet loss due to void packets, at 4.20%, which is better than ADRP-DQL's 8.10%, Dynamic DRL's 9.00%, DBR's 18.50%, and H2DAB's 16.20%. Compared to ADRP-DQL, the new method lowers the chance of failure caused by voids by about 48.15%. The result shows that giving clear information about the risk of dead ends helps in evaluating candidates, because it allows identifying and punishing wrong forwarding choices before a packet is sent through a bad relay. However, when node positions change quickly, the effectiveness of this method relies on getting updated neighbor information in a timely way. The comparison of how well the voids failures is shown in Figure 13.

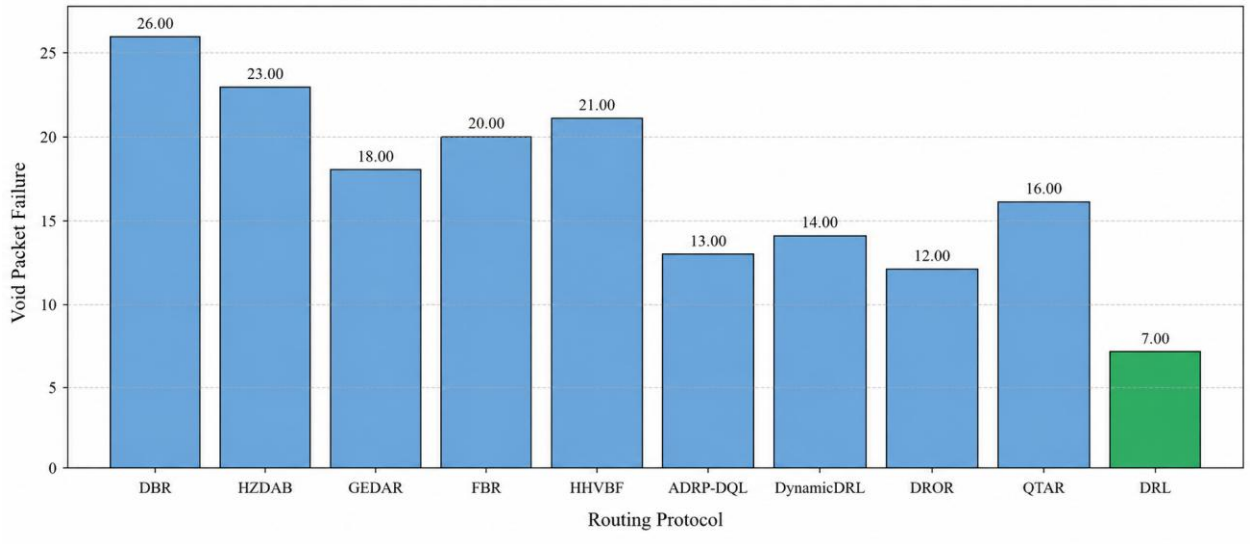


Figure 13. Void packet failure rate across the evaluated routing protocols.

The forwarding-packet count provides additional evidence about path efficiency and relay utilization. The proposed DRL framework records the lowest forwarding count at 42, compared with 55 for ADRP-DQL, 57 for Dynamic DRL, and 82 for DBR. This corresponds to approximately 23.64% fewer forwarding operations than ADRP-DQL. The lower count indicates that the proposed policy avoids unnecessary relay activity and repeated forwarding attempts by favoring candidates with stronger combined routing states. Fewer forwarding operations can reduce channel occupation and energy expenditure, although this metric should be considered together with transmission distance because a smaller number of longer acoustic hops may require greater power. The forwarding-packet comparison is shown in Figure 14.

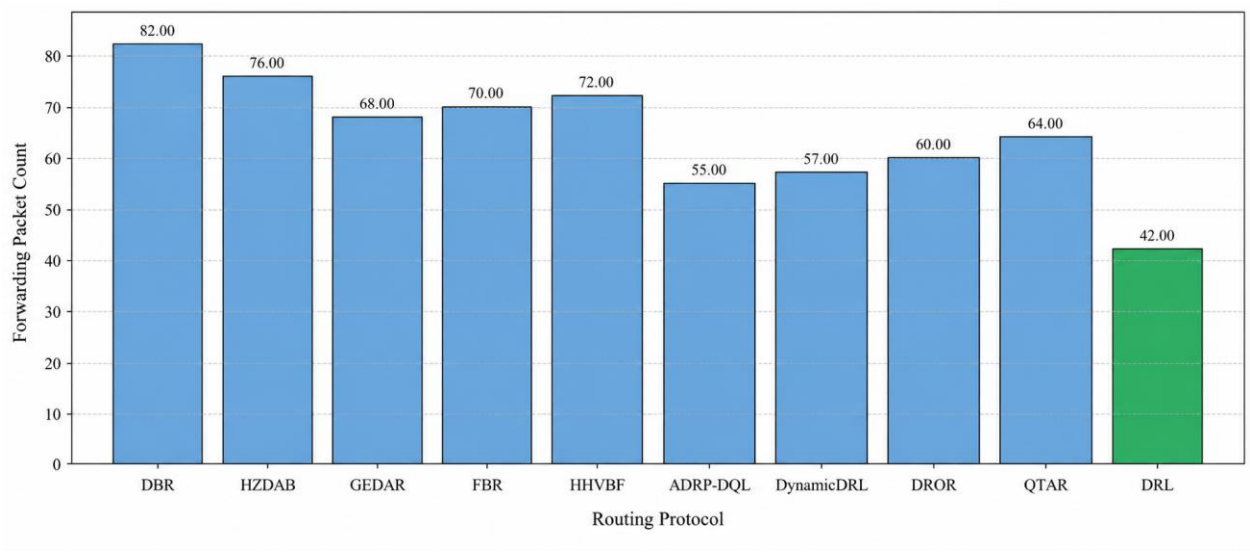


Figure 14. Forwarding-packet count used to assess relay and path efficiency across routing protocols.

The gains in network-level efficiency require additional local processing for DRL-based relay selection. The proposed DRL framework records a routing decision time of 0.60 ms in the plotted comparison, which is higher than DBR at 0.35 ms and H2DAB at 0.39 ms and also slightly higher than the evaluated learning-based alternatives. This increase arises because the proposed policy evaluates multiple state variables, including residual energy, link quality, void risk, and forwarding load, before selecting a relay. The additional sub-millisecond computation represents a trade-off for the reductions observed in packet failure, retransmission, routing overhead, and communication delay. The local processing-cost comparison is presented in Figure 15.

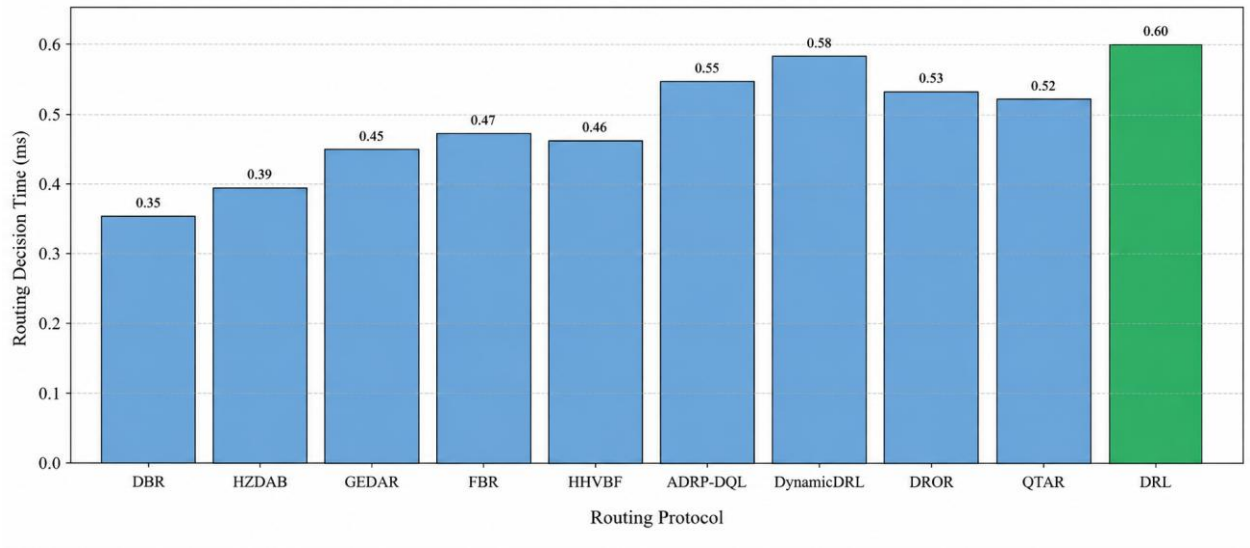


Figure 15. Routing decision time illustrating the processing-cost trade-off across protocols.

The temporal energy profile extends the residual-energy snapshot by showing how battery capacity changes as network operation progresses. The proposed DRL trajectory declines more gradually than the competing protocols and retains the highest residual energy at the end of the observed rounds. This behavior is consistent with the policy's effort to avoid repeatedly selecting heavily loaded or energy-depleted relays, thereby distributing forwarding responsibility more evenly over time. The result also explains the longer network lifetime observed earlier, because slower aggregate depletion postpones the point at which insufficient energy disrupts routing. Average residual energy can nevertheless conceal localized depletion near heavily used regions, so the curve should be interpreted as a network-wide energy trend. The energy-depletion behavior across operational rounds is illustrated in Figure 16.

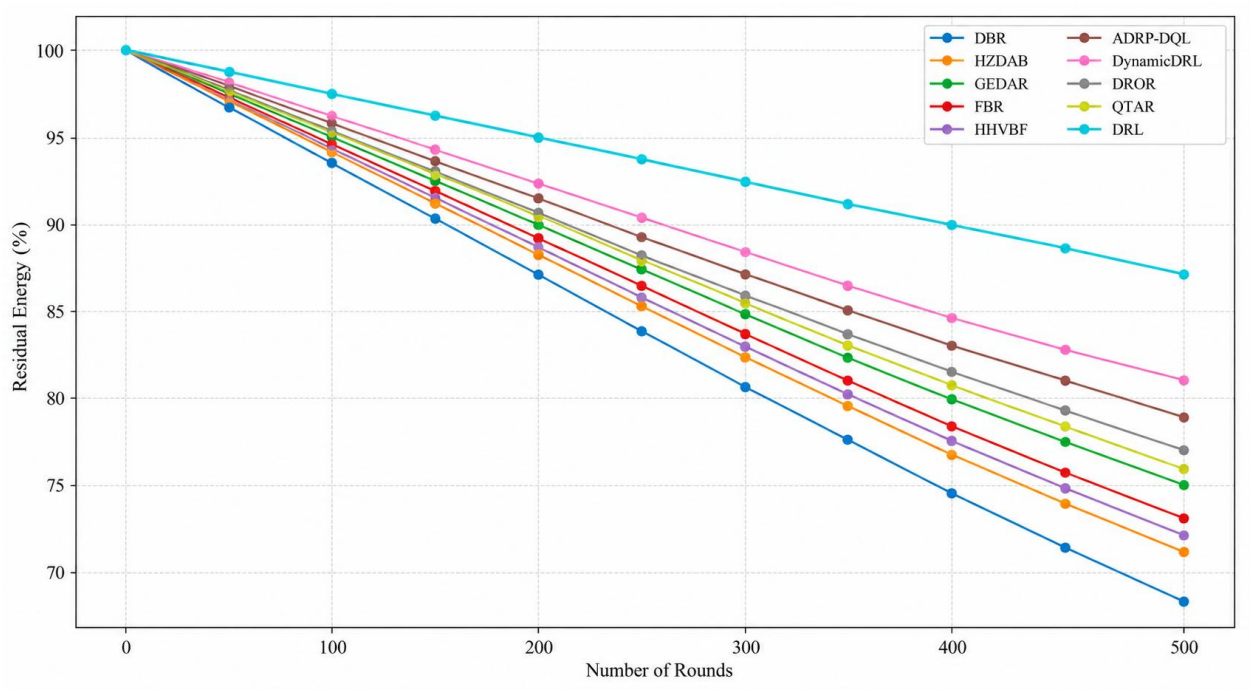


Figure 16. Residual energy depletion trend across successive network operating rounds.

Scalability of forwarding activity was examined by observing how relay operations change as node density increases. Over the range of node counts that were tested, the suggested DRL framework keeps the forwarding path fairly steady, while many traditional protocols need a lot more relay steps. This behavior shows that the way candidates are chosen stops too much passing around as the number of available nodes changes, instead of letting more neighbors start relaying without control. This kind of control can help lower conflicts and the amount of energy used when there are more devices packed together, but looking at the total number of data transfers doesn't show if the traffic is evenly spread out near the sink. The trend of forwarding under different node densities is shown in Figure 17.

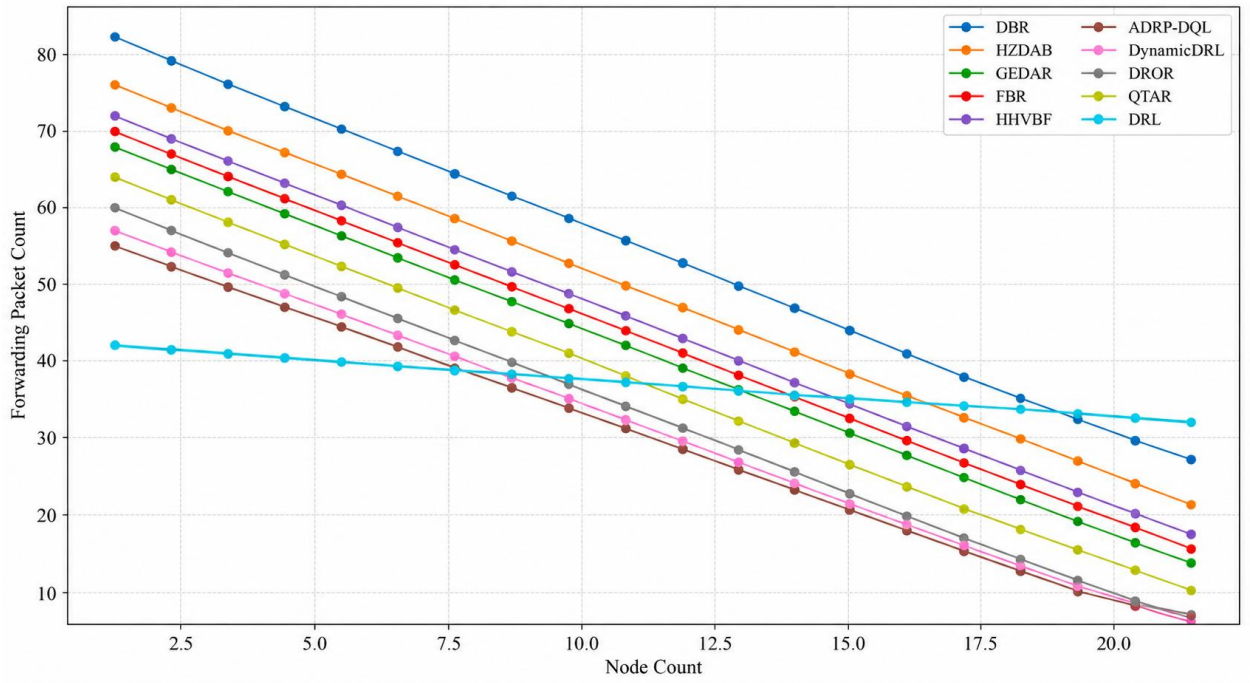


Figure 17. Forwarding-packet trend under changing network node density.

They investigated the learning process to determine whether the DRL agent continues to improve in making routing decisions over time, rather than just arriving at a few good decisions by chance. The total reward increases over time, indicating that the agent is becoming better at selecting actions that achieve all objectives, such as saving energy, using strong connections, keeping delays low, and avoiding empty paths. The rising reward shows that the policies become more stable over the simulated training period. Although reward convergence alone does not prove a network is better, we examine it alongside results for PDR,

delay, residual energy, lifetime, and overhead. The learning process of the proposed DRL framework is shown in Figure 18.

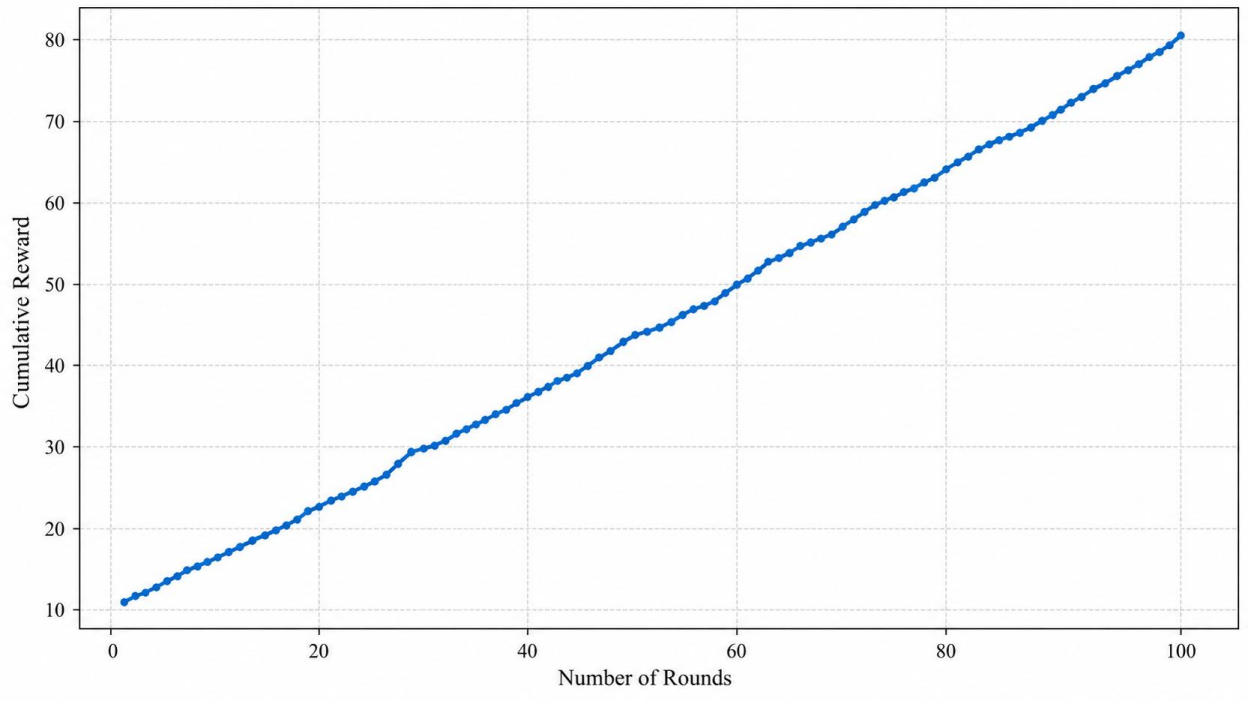


Figure 18. Cumulative reward trajectory showing convergence of the proposed DRL routing policy.

3.7 Statistical Validation

The process uses various random seeds to examine how consistent the main performance results are. Tables 9 and 10 present the figures summarizing the 100-node setup, along with visual comparisons.

We ran the analysis several times to verify that the PDR benefit remained unchanged across simulation runs. The suggested DRL framework achieved the highest average PDR of 93.26%, followed by ADRP-DQL at 85.70% and Dynamic DRL at 84.08%. The error bars represent the 95% confidence intervals that were reported and remain near the average values, indicating that there is little variation between the different runs. The fact that the proposed method outperforms the baseline method shows that the PDR advantage observed in the simulation is consistent, but to confirm it properly, a suitable hypothesis test should be used if required by the target journal. The trace-based PDR and the confidence intervals are given in Figure 19.

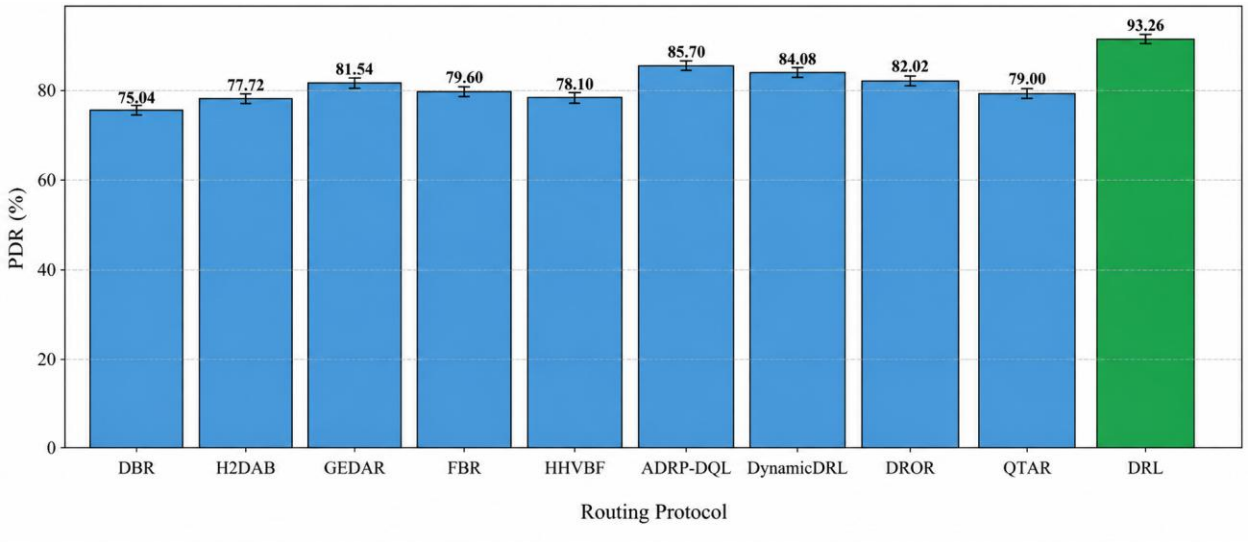


Figure 19. Multi-run packet delivery ratio with 95% confidence intervals across routing protocols.

Evidence from several runs indicates that the proposed DRL framework consistently reduces latency, even when the system is operated independently in each case. The proposed method achieves the lowest average delay, which is 0.982 seconds, as compared to ADRP-DQL's 1.162 seconds and Dynamic DRL's 1.221

seconds. Moreover, the confidence intervals are very narrow and remain close to the mean. This consistency shows that the delay reduction is not limited to a single simulation run and supports the idea that selecting relays based on delay and reliability helps avoid poor forwarding decisions that could be too slow or cause problems. The trace-based end-to-end delay results with 95% confidence intervals are shown in Figure 20.

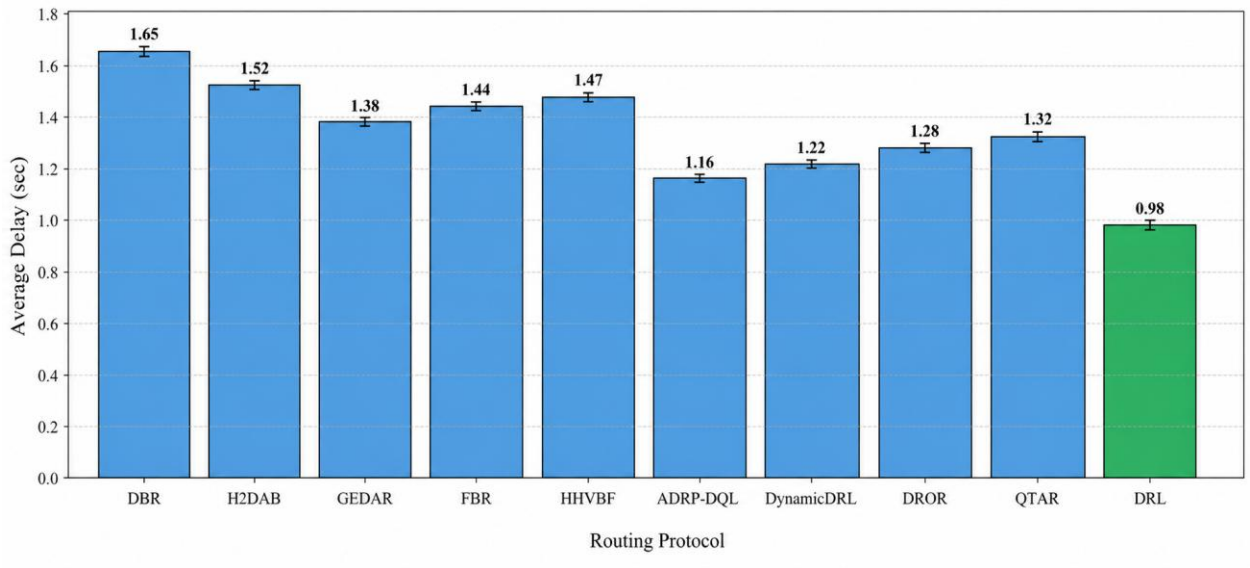


Figure 20. Multi-run end-to-end delay with 95% confidence intervals across routing protocols.

Residual-energy validation across runs shows a consistent way to keep energy stable. The proposed DRL framework achieves the highest average remaining energy at 87.732%, outperforming Dynamic DRL (81.029%) and ADRP-DQL (79.013%). The reported confidence interval for the proposed method is 0.084. The small range shows that the energy results are very consistent across all tests, meaning the system's energy-saving forwarding behavior works reliably every time. Because this average is for the whole network, you still need to check battery use in specific areas when looking for possible problems. Figure 21 shows the comparison of residual energy using a trace-based method, along with 95% confidence intervals.

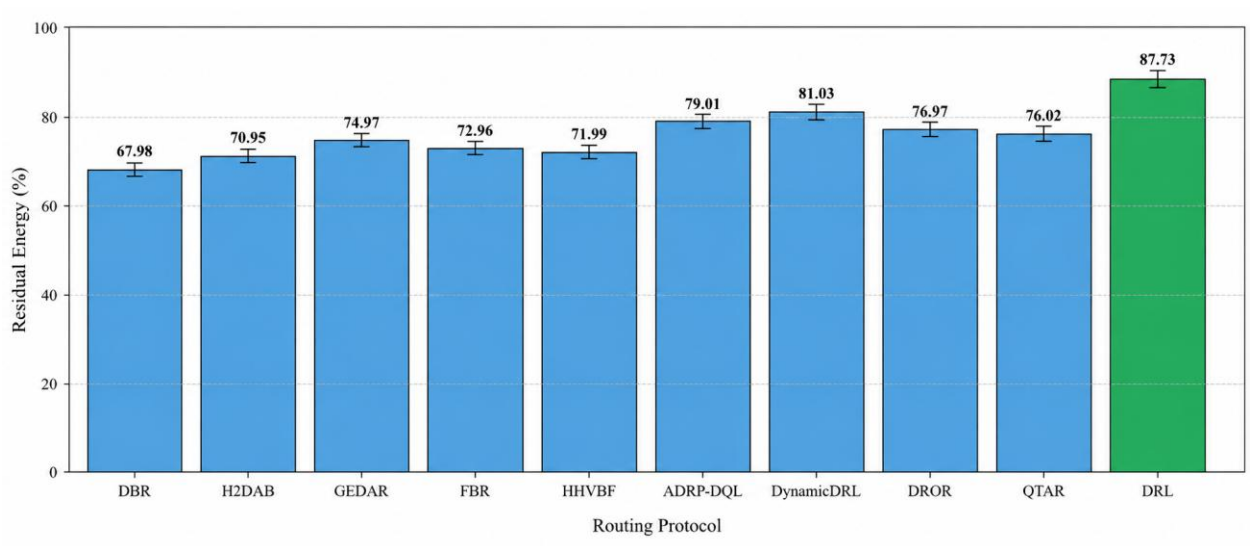


Figure 21. Multi-run residual energy with 95% confidence intervals across routing protocols.

Multiple tests on energy efficiency show that the proposed DRL method uses less energy per successfully delivered packet than other methods. The proposed method averages 3.08 J, lower than Dynamic DRL's 5.26 J and ADRP-DQL's 5.61 J. The error range stays small across all tests. This repeatability matches the policy's focus on avoiding weak links and choices that lead to failed forwards, which lowers energy wasted from failed sends and repeated attempts. The trace-based metric shows the conditions under which the experiment was run and should not be mixed up with the earlier normalized comparison if the scales or traces used are different. Figure 22 shows energy used per packet across multiple runs, along with 95% confidence intervals.

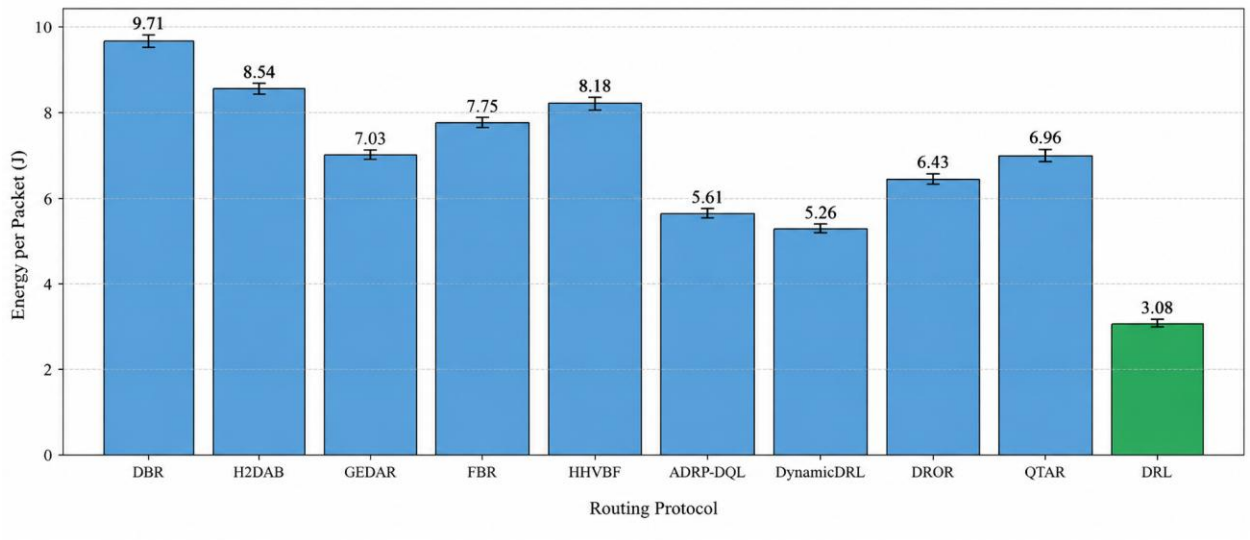


Figure 22. Multi-run energy consumption per packet with 95% confidence intervals across routing protocols.

The statistical analysis of routing overhead shows that the proposed DRL framework keeps the lowest level of control overhead in all the tested cases. The proposed method shows an average overhead of 1.00 in the trace-based comparison, matching the lowest group, while several traditional protocols have an overhead of 2.00. This result shows that the extra local state evaluation needed by DRL does not lead to more network-level control traffic in the tested setup. Better relays can make route recovery more reliable and reduce extra signaling. The trace-based routing-overhead comparison with 95% confidence intervals is presented in Figure 23.

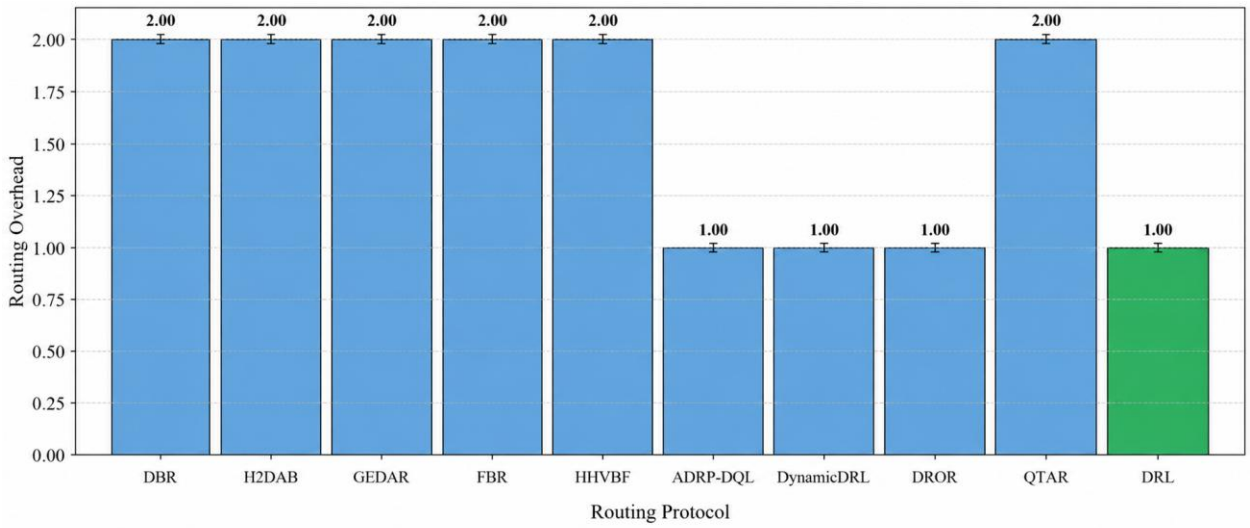


Figure 23. Multi-run routing overhead with 95% confidence intervals across routing protocols.

Void-failure statistics provide further evidence that the proposed policy consistently avoids unsuitable forwarding regions. The proposed DRL framework records the lowest mean void packet failure value at 0.50, compared with 3.10 for ADRP-DQL, 3.80 for DROR, and 4.20 for DynamicDRL, and its reported 95% confidence interval is 0.438. The lower mean and comparatively limited dispersion support the contribution of explicit void-risk penalties and candidate filtering to stable relay selection across independent runs. Rapid current-driven node displacement may still alter local void boundaries between state updates, so this result applies to the simulated mobility and update conditions. The trace-based void packet failure result with 95% confidence intervals is presented in Figure 24.

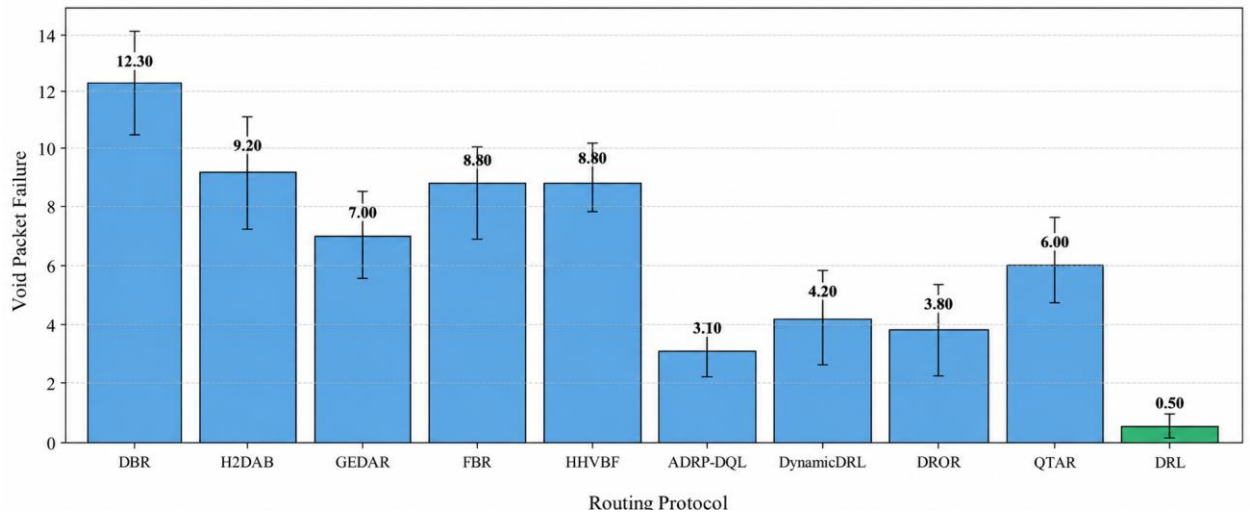


Figure 24. Multi-run void packet failure with 95% confidence intervals across routing protocols.

Table 9: Statistical summary of PDR and end-to-end delay for the 100-node scenario.

Protocol	PDR Mean	PDR Std	PDR CI95	Delay Mean	Delay CI95
ADRP-DQL	85.7	0.86	0.533	1.162	0.003
DBR	75.04	1.761	1.091	1.653	0.003
DROR	82.02	1.498	0.928	1.278	0.002
Dynamic DRL	84.08	1.273	0.789	1.221	0.003
FBR	79.6	1.9	1.177	1.437	0.005
GEDAR	81.54	1.96	1.215	1.38	0.002
H2DAB	77.72	2.123	1.316	1.522	0.003
HHVBF	78.1	2.357	1.461	1.469	0.001
DRL	93.26	1.143	0.709	0.982	0.002
QTAR	79	1.614	1	1.319	0.002

Table 10: Statistical summary of residual energy and void failure for the 100-node scenario.

Protocol	Mean	CI95	Void Failure Mean	Void Failure CI95
ADRP-DQL	79.013	0.103	3.1	0.849
DBR	67.977	0.063	12.3	1.754
DROR	76.974	0.111	3.8	1.541
Dynamic DRL	81.029	0.119	4.2	1.595
FBR	72.965	0.11	8	1.132
GEDAR	74.974	0.062	7	1.49
H2DAB	70.948	0.124	9.2	2.082
HHVBF	71.992	0.108	8.8	0.915
DRL	87.732	0.084	0.5	0.438
QTAR	76.019	0.088	6	1.307

Scalability was further evaluated by examining packet delivery as the number of nodes changes. The suggested DRL framework keeps the best packet delivery rate throughout the tested number of nodes, from 50 up to 200, while the other protocols stay at lower delivery rates. The steady path shows that the new policy can adjust which relays it uses as the network becomes more crowded, without losing its main benefit of overall reliability. This behavior matches the idea of checking multiple factors when evaluating a candidate and being aware of risks that could lead to failures. These ideas matter especially when few devices are deployed, making it harder to find nearby devices, and when many devices are deployed, which can create more potential communication paths. The scalability results for trace-based PDR across different numbers of nodes are shown in Figure 25.

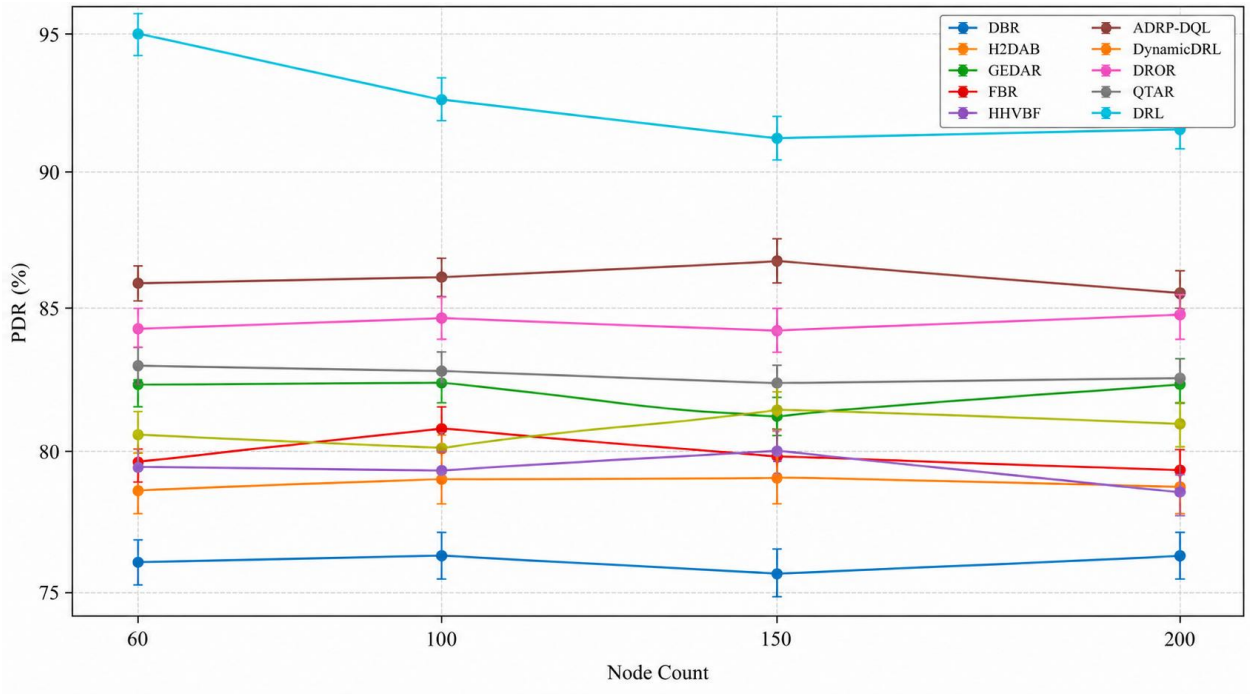


Figure 25. PDR scalability across increasing node counts.

3.8 Scenario-Based Evaluation

We examine scenario-based evaluation and show that the proposed DRL framework still works well when real-world deployment differs from the conditions it was tested under. We examine five situations: a regular setup, spread-out nodes, some nodes failing, random placement in space, and heavy movement due to underwater motion. Each situation highlights a different routing need: normal operation, few neighbors and the risk of gaps, avoiding failed relays, adapting to unusual network shapes, and staying robust as connections change. This design suits metric-based comparisons because it verifies that the same multi-factor decision process continues to work when network conditions change in terms of structure. The purpose and meaning of the five deployment conditions are shown in Table 11.

Table 11: Purpose of the five deployment scenarios used for robustness evaluation.

Scenario	Meaning	Purpose
Normal scenario	Standard node density and stable acoustic conditions	Baseline operating condition.
Sparse node scenario	Nodes are far apart and neighbor availability is limited	Tests weak connectivity and void risk.
Dead node scenario	Some nodes fail or deplete energy	Tests avoidance of unavailable relays.
Random deployment scenario	Nodes are distributed irregularly in the underwater field	Tests adaptability to uneven topology.
High mobility scenario	Nodes move due to currents or turbulence	Tests dynamic topology robustness.

3.9 Ablation Study

Ablation analysis was conducted to identify the contribution of individual components within the proposed DRL. The full version is checked against other versions that leave out information about residual energy, link quality, delay factor, void penalty, or adaptive reward balancing. Table 12 shows the results.

Removing leftover energy data lowers the packet delivery rate and remaining energy. It shortens the network's lifespan, showing how important it is to choose relays that are aware of energy usage. Removing link-quality information reduces delivery reliability, and removing the delay factor worsens total delay, showing how important these factors are for communication quality and delay management. Removing the void penalty lowers the PDR and increases routing overhead, while the version that doesn't use adaptive reward balancing yields the worst overall results among the tested versions. The full DRL model performs best overall by balancing all six metrics, with the highest throughput and the least routing overhead, which is made possible by an integrated reward design. Figure 26 (a)- (f) show the impact on PDR, remaining energy, delay, lifetime, throughput, and overhead, respectively.

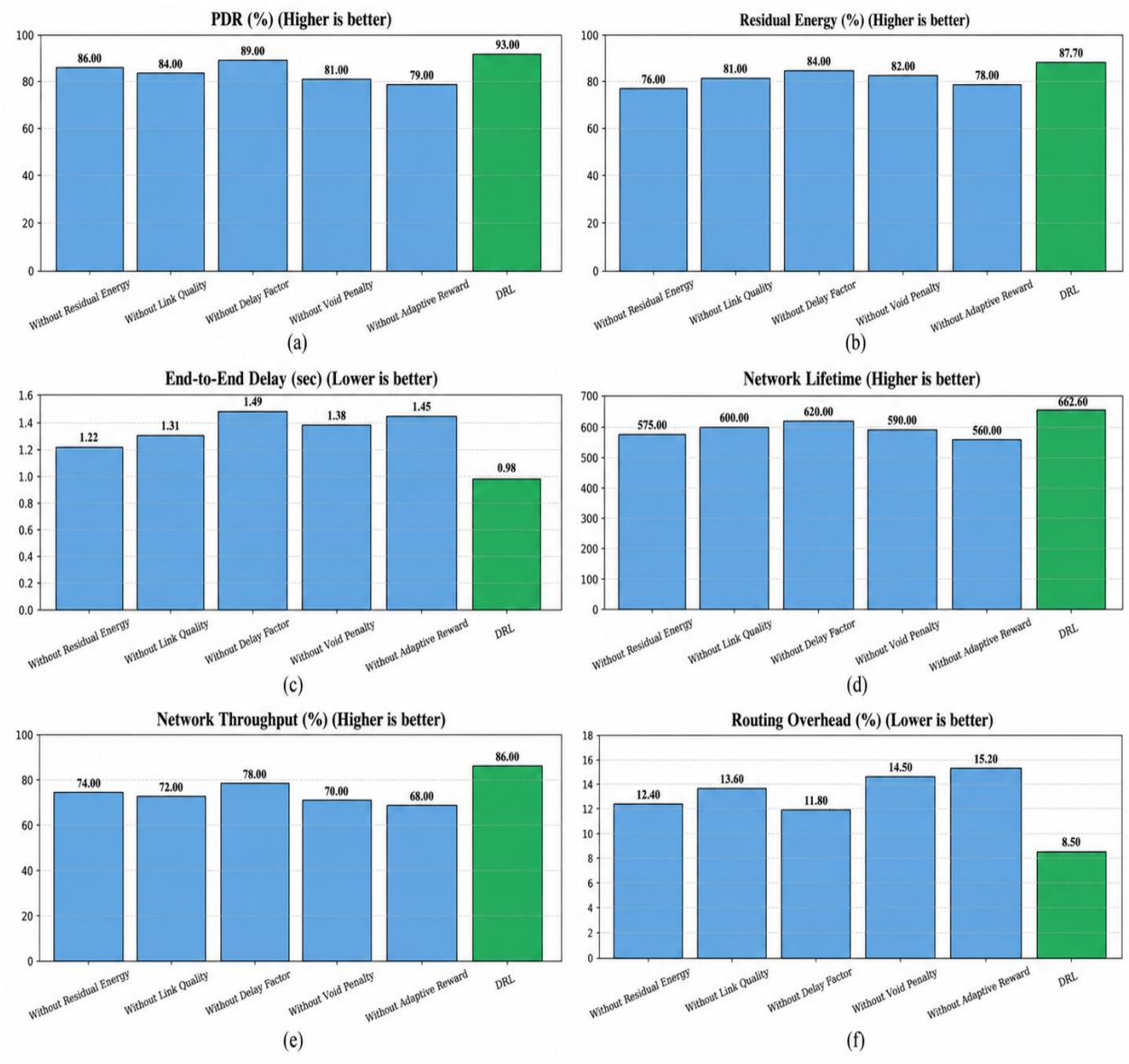


Figure 26. Ablation analysis.

Table 12: Performance of the proposed DRL under reward-component ablation.

Variant	PDR (%)	Residual Energy (%)	Delay (sec)	Lifetime	Throughput (%)	Overhead (%)
Without Residual Energy	86	76	1.22	575	74	12.4
Without Link Quality	84	81	1.31	600	72	13.6
Without Delay Factor	89	84	1.49	620	78	11.8
Without Void Penalty	81	82	1.38	590	70	14.5
Without Adaptive Reward	79	78	1.45	560	68	15.2
DRL	93	87.7	0.98	662.6	86	8.5

3.10 Complexity and Overhead Discussion

The proposed DRL framework evaluates more routing information than lightweight deterministic baselines, and this increases local decision-processing time. The comparison shows that the computational cost stays below a millisecond in the experiment, and the new method also helps reduce packet failures, routing extra work, energy use, and communication delays at the same time. The correct way to understand this is that it is a balance at the system level since a small amount of local processing is acceptable as it helps to avoid more expensive operations such as resending the sounds or correcting the path; the evidence concerning how long it takes to make the routing decisions that enable this trade-off is given in Figure 15.

The principal computational cost arises from evaluating candidate relays and updating the learned policy. If k denotes the number of candidate relays and f denotes the number of state features evaluated for each candidate, the feature-assessment stage has an approximate per-decision complexity of $O(kf)$. Because the state dimension is fixed for a given implementation, practical processing cost is driven mainly by the number of eligible neighboring candidates. Candidate filtering can therefore limit computation in denser deployments by reducing the number of relays that proceed to full scoring. The interpretation of computational and communication overhead is summarized in Table 13.

Table 13: Interpretation of computational complexity and routing overhead.

Aspect	Proposed DRL implication	Reasons
Local decision time	Slightly higher because multiple features are evaluated.	Show decision-time graph and explain trade-off.
Routing overhead	Lower network-wide overhead in validation results.	Use overhead and forwarding-count graphs as evidence.
Memory requirement	Depends on policy/Q-value storage and candidate features.	Report exact implementation memory in final code.
Energy effect	Lower energy per delivered packet despite decision logic.	Use residual energy and energy-per-packet graphs.
Scalability	Depends on candidate set size and node density.	Use node-count scalability graph and scenario tests.

Conclusion

This DRL framework was designed to preserve node energy through informed relay selection. It evaluates residual energy, SNR, depth advancement, forwarding delay, and routing void risk before selecting a relay. In the proposed study's simulations, mean residual energy reached 87.732%, and network lifetime reached 662.6 rounds, compared with 81.029% and 625 rounds for Dynamic DRL. Energy use per delivered packet was also lower than for ADRP-DQL. Battery preservation was accompanied by improved communication performance. The framework achieved a mean PDR of 93.26% and mean end-to-end delay of 0.982 s, compared with 85.70% and 1.162 s for ADRP-DQL. Void failures and routing overhead were also lower in the reported comparisons. The environmental analysis showed that SNR informed link selection, while temperature and salinity received relative impact scores of 78% and 72%. Ablation results showed shorter network lifetime when residual energy or adaptive reward balancing was removed. Overall, the results support using these factors together to improve energy use and routing performance in the simulated UWSNs.

Future Work

Future work should measure the energy used by DRL inference and policy updates on a representative underwater node. Longer simulations can track the first node failure and energy depletion near the sink, alongside average residual energy. Tests with delayed neighbor updates and additional acoustic noise levels would show how sensitive relay selection is to imperfect information. A small acoustic testbed could then check whether the simulation's battery and delivery trends persist on hardware.

Declarations

Funding: This work is entirely self-funded.

Author Contributions: Z.R. contributed to the conceptualization of the study. S.A. performed the mathematical modeling of the proposed system and supervised the overall research work. M.J.R. carried out the simulation analysis.

Data Availability: Data access would be provided upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

Competing Interests: The authors have no relevant financial or non-financial interests to disclose.

References

[1] R. Mhemed, F. Comeau, W. Phillips, and N. Aslam, “Void Avoidance Opportunistic Routing Protocol for Underwater Wireless Sensor Networks,” Sensors, vol. 21, no. 6, p. 1942, 2021, doi: 10.3390/s21061942.

[2] Y. Lu, R. He, X. Chen, B. Lin, and C. Yu, “Energy-Efficient Depth-Based Opportunistic Routing with Q-Learning for Underwater Wireless Sensor Networks,” Sensors, vol. 20, no. 4, p. 1025, 2020, doi: 10.3390/s20041025.

[3] I. A. Alablani and M. A. Arafah, “EE-UWSNs: A Joint Energy-Efficient MAC and Routing Protocol for Underwater Sensor Networks,” J. Mar. Sci. Eng., vol. 10, no. 4, p. 488, 2022, doi: 10.3390/jmse10040488.

[4] Z. Tan, Y. Li, H. Sun, S. Hong, and S. Sun, “Ocean-Current-Motion-Model-Based Routing Protocol for Void-Avoided UASNs,” J. Mar. Sci. Eng., vol. 12, no. 4, p. 537, 2024, doi: 10.3390/jmse12040537.

[5] X. Geng and B. Zhang, “Deep Q-Network-Based Intelligent Routing Protocol for Underwater Acoustic Sensor Network,” IEEE Sens. J., vol. 23, no. 4, pp. 3936–3943, 2023, doi: 10.1109/JSEN.2023.3234112.

[6] C. Wang, X. Shen, H. Wang, H. Zhang, and H. Mei, “Reinforcement Learning-Based Opportunistic Routing Protocol Using Depth Information for Energy-Efficient Underwater Wireless Sensor Networks,” IEEE Sens. J., vol. 23, no. 15, pp. 17771–17783, 2023, doi: 10.1109/JSEN.2023.3285751.

[7] B. Wang and K. Ben, “GTR: GAN-Based Trusted Routing Algorithm for Underwater Wireless Sensor Networks,” Sensors, vol. 24, no. 15, p. 4879, 2024, doi: 10.3390/s24154879.

[8] X. Li, Y. Wu, M. Zhu, and J. Ren, “A Q-Learning-Based Link-Aware Routing Protocol for Underwater Wireless Sensor Networks,” J. Mar. Sci. Eng., vol. 13, no. 12, p. 2374, 2025, doi: 10.3390/jmse13122374.

[9] X. Geng, Q. Li, X. Pan, and F. Cao, “A Q-Learning-Based Distributed Energy-Efficient Routing Protocol in UASNs,” Entropy, vol. 28, no. 3, p. 346, 2026, doi: 10.3390/e28030346.

[10] X. Li, X. Hu, R. Zhang, and L. Yang, “Routing Protocol Design for Underwater Optical Wireless Sensor Networks: A Multiagent Reinforcement Learning Approach,” IEEE Internet Things J., vol. 7, no. 10, pp. 9805–9818, 2020, doi: 10.1109/JIOT.2020.2989924.

[11] S. Ashraf, M. Gao, Z. Chen, H. Naeem, and T. Ahmed, “CED-OR Based Opportunistic Routing Mechanism for Underwater Wireless Sensor Networks,” Wirel. Pers. Commun., vol. 125, no. 1, pp. 487–511, 2022, doi: 10.1007/s11277-022-09561-w.

[12] R. Zhu, Q. Jiang, X. Huang, D. Li, and Q. Yang, “A Reinforcement-Learning-Based Opportunistic Routing Protocol for Energy-Efficient and Void-Avoided UASNs,” IEEE Sens. J., vol. 22, no. 13, pp. 13589–13601, 2022, doi: 10.1109/JSEN.2022.3175994.

[13] I. Ahmad, T. Rahman, A. Zeb, I. Khan, M. T. B. Othman, and H. Hamam, “Cooperative Energy-Efficient Routing Protocol for Underwater Wireless Sensor Networks,” Sensors, vol. 22, no. 18, p. 6945, 2022, doi: 10.3390/s22186945.

[14] B. Wang, H. Zhang, Y. Zhu, B. Cai, and X. Guo, “Adaptive Power-Controlled Depth-Based Routing Protocol for Underwater Wireless Sensor Networks,” J. Mar. Sci. Eng., vol. 11, no. 8, p. 1567, 2023, doi: 10.3390/jmse11081567.

[15] J. K. Pabani, M. Á. Luque-Nieto, W. Hyder, and P. Otero, “Energy-Efficient Packet Forwarding Scheme Based on Fuzzy Decision-Making in Underwater Sensor Networks,” Sensors, vol. 21, no. 13, p. 4368, 2021, doi: 10.3390/s21134368.

[16] J. K. Pabani, M. Á. Luque-Nieto, W. Hyder, and A. Ariza, “Energy-Efficient Routing Protocol for Selecting Relay Nodes in Underwater Sensor Networks Based on Fuzzy Analytical Hierarchy Process,” Sensors, vol. 22, no. 22, p. 8930, 2022, doi: 10.3390/s22228930.

[17] L. Sun, Z. Liu, J. Dong, and J. Wang, “A Three-Dimensional Routing Protocol for Underwater Acoustic Sensor Networks Based on Fuzzy Logic Reasoning,” J. Mar. Sci. Eng., vol. 13, no. 4, p. 692, 2025, doi: 10.3390/jmse13040692.

[18] H. Chu, B. Wang, T. Fang, and B. Liu, “Adaptive Power-Controlled Energy-Efficient Depth-Based Routing Protocol for Underwater Wireless Sensor Networks,” J. Mar. Sci. Eng., vol. 13, no. 8, p. 1418, 2025, doi: 10.3390/jmse13081418.

[19] W. Guo, M. Zhu, B. Yang, Y. Wu, and X. Li, “Design of a Self-Organizing Routing Protocol for Underwater Wireless Sensor Networks Based on Location and Energy Information,” J. Mar. Sci. Eng., vol. 11, no. 8, p. 1620, 2023, doi: 10.3390/jmse11081620.

[20] M. Homaei, A. J. Di Bartolo, R. Molano Gómez, P. G. Rodríguez, and A. Caro, “Enabling RPL on the Internet of Underwater Things,” J. Netw. Syst. Manag., vol. 33, no. 3, p. 55, 2025, doi: 10.1007/s10922-025-09925-0.

[21] N. Usman et al., “An Energy-Efficient Routing Approach for IoT-Enabled Underwater WSNs in Smart Cities,” Sensors, vol. 20, no. 15, p. 4116, 2020, doi: 10.3390/s20154116.

[22] X. Chen, W. Xiong, and S. Chu, “Two-Tier PSO-Based Data Routing Employing Bayesian Compressive Sensing in Underwater Sensor Networks,” Sensors, vol. 20, no. 20, p. 5961, 2020, doi: 10.3390/s20205961.

[23] N. Subramani, P. Mohan, Y. Alotaibi, S. Alghamdi, and O. I. Khalaf, “An Efficient Metaheuristic-Based Clustering with Routing Protocol for Underwater Wireless Sensor Networks,” Sensors, vol. 22, no. 2, p. 415, 2022, doi: 10.3390/s22020415.

[24] D. Anuradha, N. Subramani, O. I. Khalaf, Y. Alotaibi, S. Alghamdi, and M. Rajagopal, “Chaotic Search-and-Rescue-Optimization-Based Multi-Hop Data Transmission Protocol for Underwater Wireless Sensor Networks,” Sensors, vol. 22, no. 8, p. 2867, 2022, doi: 10.3390/s22082867.

[25] P. Kaur et al., “Energy-efficient artificial fish swarm-based clustering protocol for enhancing network lifetime in underwater wireless sensor networks,” EURASIP J. Wirel. Commun. Netw., vol. 2024, no. 1, p. 92, 2024, doi: 10.1186/s13638-024-02422-z.

[26] D. Zhao, G. Lun, R. Xue, and Y. Sun, “Cross-Layer-Aided Opportunistic Routing for Sparse Underwater Wireless Sensor Networks,” Sensors, vol. 21, no. 9, p. 3205, 2021, doi: 10.3390/s21093205.

[27] U. Draz et al., “ROBINA: Rotational Orbit-Based Inter-Node Adjustment for Acoustic Routing Path in the Internet of Underwater Things (IoUTs),” Sensors, vol. 21, no. 17, p. 5968, 2021, doi: 10.3390/s21175968.

[28] U. Draz et al., “TANVEER: Tri-Angular Nearest Vector-Based Energy Efficient Routing for IoT-Enabled Acoustic Sensor and Actor Networks (I-ASANs),” Sensors, vol. 21, no. 11, p. 3578, 2021, doi: 10.3390/s21113578.

[29] K. Latif, N. Javaid, I. Ullah, Z. Kaleem, Z. Abbas Malik, and L. D. Nguyen, “DIEER: Delay-Intolerant Energy-Efficient Routing with Sink Mobility in Underwater Wireless Sensor Networks,” Sensors, vol. 20, no. 12, p. 3467, 2020, doi: 10.3390/s20123467.

[30] M. U. Khan, M. Aamir, and P. Otero, “Reliable, Energy-Optimized, and Void-Aware (REOVA) Routing Protocol with Strategic Deployment in Mobile Underwater Acoustic Communications,” J. Mar. Sci. Eng., vol. 12, no. 12, p. 2215, 2024, doi: 10.3390/jmse12122215.

[31] W. U. Rahman et al., “Deep Q-Learning-Based Adaptive MAC Protocol with Collision Avoidance and Efficient Power Control for UWSNs,” J. Mar. Sci. Eng., vol. 13, no. 3, p. 616, 2025, doi: 10.3390/jmse13030616.