INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2026, VOL. 72, NO. 2, PP. 1-10
Manuscript received January 20, 2026; revised April 2026. doi: 10.24425/ijet.2026.157924

Deep Reinforcement Learning-enabled energy-efficient routing protocol for underwater wireless sensor networks

Yogeshwary Bommenahalli Huchegowda, and Mahadeva Prasad M

***Abstract*—Underwater wireless sensor networks are widely used in sea and ocean exploration, monitoring of the environment, defense surveillance. These applications are restricted by limited energy availability, propagation delay of acoustic signal, and topology changes. To address these issues, a reinforcement learning (RL)-based routing protocol that combines energy-aware clustering with Q-learning to improve packet forwarding efficiency is proposed in this paper. In this approach, the role of each autonomous agent is performed by sensor node and forwarding actions based on residual energy, hop count, and distance to the sink are adaptively selected. MATLAB simulation results demonstrate that the proposed scheme achieves a packet delivery ratio (PDR) of 95.2%. Compared with vector-based forwarding (VBF) and reinforcement learning-based opportunistic routing (RLOR), the achieved PDR is 7.6% and 3.7% higher, respectively. The improvement of performance is mainly attributed to adaptive Q-learning-based next-hop selection and energy-aware clustering, which reduce redundant and long-distance transmissions and avoid routing voids. Moreover, the proposed protocol extends network lifetime to 5000 iterations, achieving improvements of 19% and 6.4%, while reducing average energy consumption by 25.7% and 13.3% compared with VBF and RLOR.**

***Keywords*—Energy-aware Clustering; Energy Efficient Routing, Q-Learning; Reinforcement Learning; Underwater Wireless Sensor Networks**

I. INTRODUCTION

THE oceans surrounding the world today have become increasingly important for environmental monitoring and exploration, as well as disaster and surveillance. To address these needs, there arises an ever continuing and intense need for efficient underwater sensor systems. To address these problems, there emerged an emerging area known as Underwater Wireless Sensor Networks (UWSNs), which offer autonomous sensing and data collection and forwarding capabilities even for difficult underwater environments. Nonetheless, it should be noted that, unlike wireless sensor networks that communicate via RF signals, there exist special limitations within an underwater network because RF signals will be highly attenuated by water [1]. UWSNs use acoustic signals which can provide long range communications. This must face challenges such as low bandwidth requirements, high latency, and large bit error rates [1],[2].

These limitations have effects on the performance of underwater sensor networks. For example, unreliable and time-varying communication links is caused by the slow propagation speed of about 1500 m/s by the acoustic waves along with multipath effects, doppler shifts, and ambient noise [2], [3].The frequent change can be observed in network topology and stability of the link due to the ocean currents which causes the continuous mobility of the underlying sensor nodes. Energy efficiency is another basic requirement since underwater sensor nodes make use of the batteries which cannot be replaced or recharged in underwater environments [3]. So, there is need to have routing algorithms that lower the redundant transmissions, reduce long-distance forwarding, and balance energy usage to increase the lifetime of the network.

To improve the data delivery in underwater, many routing protocols have been proposed. Many traditional terrestrial routing techniques are not suitable due to the dynamic nature and scarcity of resources in underwater. The geographic routing approaches though they are simple, often suffer from routing voids, while opportunistic forwarding gives rise to redundant transmissions and uneven energy depletion [4]. In similar ways, static sink-based architecture is indeed useful to simplify the data collection procedure but still suffer during the selection of reliable relay paths given mobility, sparse deployment, and fluctuating channel characteristics. These limitations suggest there is a need for routing strategies that can intelligently adapt to environmental conditions and node states.

To overcome these issues, there has been interest in learning-based and adaptive routing approaches that have robustness and adaptability for UWSNs. Specifically, Reinforcement Learning (RL) is an approach that enables UWSN nodes to learn and make judgments about forwarding based on interactions with the surrounding environment. Based on RL, UWSN nodes can make more adaptive routing decisions compared with deterministic routing methods because they can weigh factors that affect routing, like residual energy and hop numbers. Overall, RL has adaptability and resilience and can thus be attractive for UWSN networks considering energy and dynamic constraints.

To address these challenges, a new UWSN routing strategy empowered with reinforcement of learning capabilities and improving energy efficiency and reliability within UWSNs has been introduced [5]. Based on energy efficient path-routing and adaptive decision-making, every node will be able to make

First Author is with Department of Electronics and Communication Engineering, Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, India (e-mail: yogeshwary.ec@sode-edu.in).

Second Author is with Department of Studies in Electronics, Hemagangothri, University of Mysore, Hassan, 573226, Karnataka, India (e-mail: prasada9@gmail.com).

autonomous optimal forwarding decisions based on local knowledge. By giving more importance to energy savings, eliminating redundant paths, and spreading data reliably towards the sink node, various disadvantages created within traditional routing methods will be eliminated. Also, based on learning capabilities, it will be possible for the technique to learn and adapt itself with acoustic channel changes.

It is crucial to consider the common network architecture employed in UWSNs to gain a better understanding of how routing decisions spread in underwater environments. The single-sink architecture, seen in figure. 1, is one of the most widely used architectures.

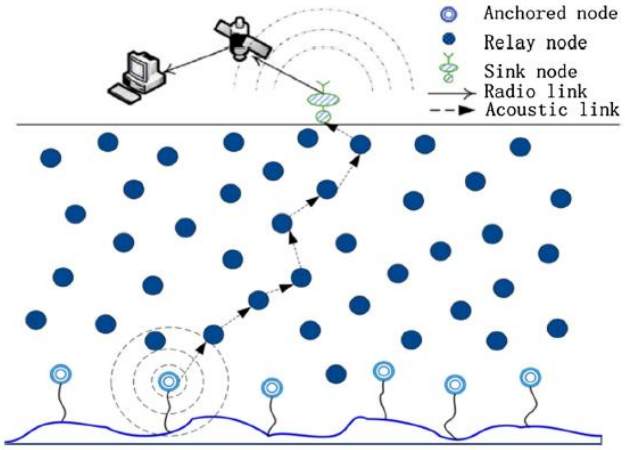

Fig. 1. Underwater sensor node deployment with acoustic multi hop routing

A single-sink architecture that comprises anchored nodes, relay nodes, and one sink node is depicted in Fig. 1. Sensing and data gathering are the main functions of the anchored nodes, which are placed on the ocean floor, whereas relay nodes are placed along various depths to search for the next route to a destination and to send the gathered packet toward the new node and gather data from the environment. In terms of routing, the sink node has both radio and acoustic modems and is the destination node. When communicating underwater, an acoustic link is utilized, and when communicating outside the water surface, a radio link is employed. Data are sensed by anchored nodes and then transmitted via relay nodes to the surface sink node. Using a radio link, the sink node delivers the data it has received from anchored nodes to a satellite, which then relays it to the shore-based control center.

The aim of the proposed work is based on developing an efficient routing scheme that tackles the typical challenges associated with underwater communications, such as high propagation delay, bandwidth constraints, energy dissipation, and dynamic network topological variations. The proposed solution tackles these challenges through adaptive, energy-efficient, and delay-aware routing. The remaining parts of this paper are organized as: the objectives are highlighted in Section II. Section III describes the related work in this area of research. Section IV presents an RL-based methodology. Section V describes the simulation results and discussions, while Section concludes the paper with key findings and future research directions.

II. OBJECTIVES

The goal of the research is to determine an energy-efficient and reliable routing protocol for underwater wireless sensor networks by integrating reinforcement learning with energy-aware clustering. The proposed algorithm helps every sensor node act like an autonomous agent that chooses forwarding paths adaptively concerning residual energy, hop count, and distance to the sink to reduce redundant transmission and long-distance transmission. The main idea of this approach is to avoid routing voids and balance energy consumption across the network for enhancing packet delivery ratio and prolonging network lifetime. The proposed routing protocol is evaluated using MATLAB simulations with practical performance metrics such as packet delivery ratio, average energy consumption, lifetime of the network, and the number of dead nodes, and compared with previous protocols including Vector-Based Forwarding (VBF) and Reinforcement Learning Opportunistic Routing (RLOR).

III. RELATED WORK

This section highlights the status of routing solutions in underwater acoustic sensor networks and corresponding studies on the use of RL in handling the challenges of UWSNs related to energy constriction, high delay in propagating signals through water, or dynamic environments. Haitao Yu et al. in [6] proved the usefulness of adaptive RL-based routing protocols in promoting reliability and efficiency over UWSN networks by considering the real-time observation of transmission distance or channel status, as well as representing the capability of nodes in the networks. This capability enhances the lifetime of the networks and the efficiency of communication in fields related to disaster response systems or oceanic explorations. Expanding these concepts, authors in [7] used the multi-agent reinforcement learning (MARL) protocol in the field of underwater optical sensor networks through the collaboration of nodes in optimizing the routing protocol in consideration of link quality or the remaining capabilities of the nodes. Another research on the use of the specifications of multimodal communication offered by the potential of networks initiated the development of the RL-based routing protocol by authors in [8], while the MARLIN protocol [9] took the initiative in the use of the details of the networks through the integration of the relay nodes or the transmission frequency in the use of the reinforcement learning-based protocols in UWSN networks. Another UWSN-related congestion reduction protocol related to delayed convergence and the potential of packet collisions in mobile networks initiated the development of the PDDQN-HHVBF through Chen et al. in [10].

Energy efficiency and adaptive learning have been taken as key considerations in the following studies. The RLBEEP algorithm [11] blended reinforcement learning, sleeping schedule, and controlled data transmission to lower energy and prolong network life. Likewise, Kumar et al. [12] focused on the dynamic routing in wireless sensor networks using Q-learning, optimizing routing paths using energy-efficient reward systems. More complex models of learning have appeared in literature as well, such as the multiattention actor double critic (MADC) framework by Wang et al. [13] for a scaled-up industry-scale IoT network. Another work on opportunistic routing and RL was done in RLOR algorithm [14] that considers the density of

the neighborhood, node depth, and residual energy of nodes. Furthermore, security-oriented RL was considered by Hajar et al. [15] in designing a multi-agent network that employed a unique reward function to handle packet-dropping attacks such as black hole and sink hole attacks. Later, Prabhu et al. [16] furthered the fact that a network of RL agents can facilitate energy-efficient routing without the aid of pre-existing routing tables. Their framework is founded on the RL framework conceptualized by Sutton and Barto [17].

On the other hand, non-learning strategies and hybrid designs remain in use in UWSN routing. The Depth-Based Routing (DBR) protocol proposed by Yan et al. in [18] is one of the pioneering geographic routing protocols for UWSNs. Although this protocol minimizes the cost of routing, it has experienced redundant deliveries and voids on the network. Additionally, a thorough overview of the routing protocols in UWSNs was introduced by M Ayaz et al. in [19], discussing RL-based solutions for UWSNs' energy efficiency, delay optimization, and topology construction. Later, adaptive energy efficient routing protocol was introduced in [20], deep reinforcement learning based multiple access for underwater acoustic networks is presented [21]. Thorough discussion on UWSN-related energy-efficient routing protocols was presented in [22] by S U Khan and adaptive clustering routing protocol was introduced by Y Sun et al. in [23], Furthermore, channel modeling and simulation frameworks for UWSNs were introduced in [24] by Chitre et al. and in [25] by R Sharma. Recently, UWSN routing has incorporated concepts from both trust management and Deep Learning for enhanced performance. For example, in [26], RL-based secure UWSN routing was introduced. Additionally, Deep RL based UWSN routing was introduced in [27] and multi hop routing with clustering in [28]. On the other hand, [29] introduced trust based secure routing design, while deep RL based adaptive modulation for acoustic communication [30] was introduced.

Although significant work has been conducted in this area, high computational complexities and low convergence rates as well as responsiveness to dynamic topological changes are still some of the major drawbacks. This incorporated work proposes an fully distributed Q-learning-based UWSN approach with adaptive reward assessment for joint optimization of UWSN-related delays, reliability, as well as energy efficiency.

IV. DESIGN METHODOLOGY

The proposed methodology combines energy-efficient clustering and Q-learning based reinforcement learning routing for underwater wireless sensor networks. The proposed methodology accurately models the underwater wireless communication channel by initially incorporating a proper acoustic propagation model for signal attenuation, propagation delay, and channel variations. The major advantages of the hybrid approach are energy efficiency in well-balanced clustering, reduced end-to-end delay due to adaptive routing, improved reliability in dynamic channels, fully distributed and online processing, and robustness in highly variable network topologies. The nodes are then grouped into energy-aware clusters; reclustering is periodic to even out energy consumption throughout the network and prolong the lifetime of the network.

The routing strategy that is used inside every cluster is based on Q-learning which is implemented in a localized manner. The optimal forwarding paths sensor nodes are selected by observing each packet's residual energy, link reliability, and topology updates. To enhance packet delivery ratio, minimize latency, and to improve energy efficiency, reward function is designed. Efficient data delivery is achieved by the combination of adaptive clustering and intelligent routing for realizing efficient and distributed decision making within the underwater environment.

4.1 Propagation and Energy Consumption Model

UWSN uses acoustic signals for communication which are significantly attenuated by environmental factors. To model this attenuation and to estimate energy consumption, both energy consumption and propagation loss components are considered.

4.1.1 Acoustic Signal Attenuation

The total attenuation A (d, f) of an acoustic signal is modeled by using both spreading loss and absorption loss over a distance 'd' at frequency 'f'. This is given as follows:

$$A\ (d, \mathrm{f}) = \boldsymbol{d^{\eta}\cdot\alpha(f)^{d}} \tag{1}$$

where η is the spreading factor is (typically 1.5), α(f) is the absorption coefficient dependent on the transmission frequency, and d is the distance between the transmitter and receiver.

This formulation is consistent with underwater acoustic modeling. [6]

The Euclidean distance 'd_{ij}' for two nodes with coordinates $(x_i\ y_i\ z_i)$ and $(x_j\ y_j\ z_j)$ is given by

$$d_{ij}=\sqrt{(x_j - x_i)^2\ +\ (y_j - y_i)^2\ +(z_j - z_i)^2} \tag{2}$$

The transmission costs and routing paths in the network are determined by using the distance metric. [6, 7]

4.1.2 Node energy consumption model

For communication and data processing, total energy used by the sensor node is computed by using the transmitted energy, received energy and data aggregation.

The transmission energy (E_{Tx}) required to send $k-$ bit data packets over a distance is computed via the following expression:

$$E_{Tx} = E_{elec}.\,k + E_{amp}\,.\,k\cdot\,d^2 \tag{3}$$

where E_{amp} is the energy required by the power amplifier according to the distance squared model and where E_{elec} is the energy used per bit to operate the transmitter circuitry [15].

The reception energy (E_{Rx}) required to receive the $k-$ bit packet [1] is given by

$$E_{Rx} = E_{elec}.\,k \tag{4}$$

To process and merge the received data, a cluster head or relay node uses energy during the data aggregation process. The data aggregation energy (E_{DA}), is given by

$$E_{DA} = E_{agg}.\,k \tag{5}$$

where E_{agg} denotes the energy required per bit for data processing.
The total energy used by a sensor node considering one cycle of transmission energy, reception energy, and aggregation energy, is

$$E_t = E_{Tx} + E_{Rx} + E_{DA} \quad (6)$$

4.2 System Architecture and Functional Overview

Cluster-based communication architecture driven by RL is implemented by addressing the issues of limited bandwidth, dynamic topologies, high propagation delay, and constrained energy resources [21,22]. The main aim is energy-aware routing that adjusts to changes in the underwater environment, including node mobility and ocean currents.

Based on residual energy levels and physical closeness, clusters are formed by grouping nodes. The cluster head is elected by cluster based on parameters such as distance to the sink, node distance, and availability of energy which helps to minimize energy consumption.

The data from the member nodes is collected by the cluster head and uses multihop communication for the data transfer to the sink. The Q-learning algorithm is used to select the next hop from its nearby nodes. Each node can gradually learn and modify its routing behavior by calculating and updating the Q values.

The reinforcement learning agent uses reward function to evaluate routing actions. Transmission is rewarded if the packets are transmitted successfully and energy efficient forwarding. Penalties are given for packet drop. Q values are updated with these rewards for the continuous improvement in the routing. The intelligent routing decisions are made based on local information and environmental input which promotes scalability and robustness.

4.3 Clustering and CH Selection

The proposed energy-efficient routing technique for UWSNs mainly depends on clustering to reduce energy consumption and enhance network lifetime. The cluster head (CH) is responsible for data forwarding and local communication within each cluster.

4.3.1 Cluster Formation

Nodes with higher energy is selected by considering residual energy to enhance network lifetime. The intracluster communication efficiency is enhanced by the nodes which are close to one another which are also used to reduce transmission energy use. The clustering process is used to balance the node mobility and the depletion of energy.

4.3.2 CH Selection

The cluster head selection is done by its Q value, which plays important role in coordination and data forwarding. The Q value is computed as follows: [13,24]

$$\boldsymbol{Q_i} = \boldsymbol{w_1}\cdot \left(\frac{\boldsymbol{E_i}}{\boldsymbol{E_0}}\right) + \boldsymbol{w_2}\cdot \left(\frac{\boldsymbol{1}}{(\boldsymbol{h_i}+\boldsymbol{1})}\right) + \boldsymbol{w_3}\cdot \boldsymbol{Degree_i} + \boldsymbol{w_4}\cdot \boldsymbol{SuccessRate_i} \quad (7)$$

Where E_i = Residual energy of node *i*, E_0 = Initial energy of the node, h_i = Hop count from node *i* to the sink, $Degree_i$ = Number of 1-hop neighbors (connectivity index), $SuccessRate_i$ = Historical packet transmission success rate of node *i* and w_1, w_2, w_3, w_4 = Weighting factors such that $w_1 + w_2 + w_3 + w_4 = 1$

The node having the highest Q-value is selected as the CH for that round.

4.3.3 Role of CHs in the Network

The CHs have two main responsibilities after being elected:

1. Data aggregation: To remove redundancy, each CH gathers sensed data from its member nodes and uses in-network data fusion. This procedure saves node energy and lowers the overall number of packets sent.

2. Multihop Routing: Using path selection based on reinforcement learning, the combined data are subsequently sent to the sink node via nearby CHs. To ensure energy efficiency and dependable delivery, next-hop CHs are selected adaptively by assessing the Q values, residual energy, and hop distance.

4.4 Reinforcement Learning-Based Routing Mechanism

Owing to their dynamic and resource-constrained characteristics, UWSNs require routing algorithms that are both adaptive and energy efficient. In the proposed method, distributed decision-making is made possible via RL, in which every sensor node independently discovers the best routing strategy through interaction with its immediate surroundings. This process enhances resilience in dynamic undersea environments and lessens reliance on global information.

4.4.1RL Framework Components

In the proposed routing method, the reinforcement learning paradigm is tailored to the operational environment of UWSNs. Every sensor node within the network operates as an independent agent that can make localized routing choices.

The state of every agent is contained within a vector that holds pertinent local information, such as the residual energy of the node, the hop distance to the sink, the Euclidean distance from neighboring nodes, and its own present role in the network (e.g., member node or cluster head). Based on this state, the node takes an applicable action from a set of predefined options such as transmitting a packet to a target neighbor, designating itself as a cluster head, or modifying its transmission power.

As soon as the agent takes an action, it receives a reward that quantitatively encodes the result of its action. This reward function is designed to encourage behavior toward increasing energy efficiency and reliability of communication. For example, successful delivery of a packet and reducing transmission distances provide rewards, whereas packet loss or high energy expenditure incurs penalties. Through this reward-driven feedback method, the agent can utilize Q-learning to update its knowledge and continuously improve its decision-making process over time.

By identifying basic elements such as the environment, state, agent, action, and reward within the context of UWSNs, the proposed framework enables intelligent and adaptive routing that conforms to the challenges of underwater communication, such as scarce energy resources, variable link quality, and frequent topology fluctuations.

4.4.2 Mathematical Modeling of RL-Based Routing

To explicitly transfer the energy and physical restrictions of UWSNs into the RL framework, a mathematical model is presented here. As a result, the learning process is guaranteed to be based on quantifiable network factors such as residual energy, distance, hop count, and reliability rather than being abstract.

In a Q-learning environment, every node functions as an agent and decides how to route based on observations made locally. The agent chooses an action after interacting with the environment, obtains a reward based on the result, and modifies its Q value to improve decision-making in the future. In the reinforcement learning paradigm, each node's state space is determined by a collection of attributes that represent its network structure and local status. These consist of the node's current role identification as either a member node or a CH, its hop count to the sink node, its residual energy level, and the Euclidean distance to nearby nodes.

Each sensor node acts as an independent agent with the ability to choose actions from a predetermined set in the suggested reinforcement learning framework. According to its present state and the surrounding circumstances, it can choose to save energy by going into sleep mode, serving as a CH, or sending data to a nearby node. The RL agent is guided toward dependable and energy-efficient routing decisions by the reward function. In addition to successful data delivery, residual energy, transmission distance, and penalties for unfavorable results are considered. The reward function is given as follows [13,15,24].

$$r_t = w_1 \cdot \left(E_{residual} / E_0\right) + w_2 \cdot \left(1 / (d + \varepsilon)\right) + w_3 \cdot \text{Success\_Flag} - w_4 \cdot \text{Penalty} \quad (8)$$

Where $E_{residual}$ is the current energy level of the node, E_0 is the initial energy, d represents the Euclidean distance to the selected neighbor or the sink and ε is a small constant to avoid division by zero, Success_Flag is set to 1 if the data transmission is successful; otherwise, it is 0.

The penalty accounts for failed transmissions or excessive delays, and w_1, w_2, w_3, and w_4 are weighting factors that control the influence of each term Q values are updated via the Bellman equation; [8],[13],[15]

$$Q_{new}(s_t, a_t) \leftarrow Q_{old}(s_t, a_t) + \alpha \left[r_t + \gamma \max Q(s_{t+1}, a) - Q_{old}(s_t, a_t)\right] \quad (9)$$

Where α: learning rate, γ: discount factor, s_t: current state, a_t: action taken, r_t : immediate reward and s_{t+1}: next state

4.5 Multihop Data Transmission and Reclustering

After the CHs are selected via a Q value-based selection mechanism, intracluster communication starts with member nodes sending their sensed data to their respective CHs. These CHs aggregate data and start multihop intercluster communication to send the aggregated data to the surface sink node.

During this stage, every CH chooses its next-hop CH through the Q-learning-based decision process presented in Section 4.4. The choice is performed by comparing the Q values of neighboring CHs and selecting the highest value, which indicates the best trade-off between residual energy, distance of transmission, and probability of successful delivery. The dynamic learning capability enables CHs to adjust their forwarding behavior adaptively based on historical performance feedback and network status.

To ensure routing efficiency in the long term, the protocol supports a reclustering process. When the residual energy of a CH falls below a specified level, it is disqualified to serve as aforwarder or cluster head. A reclustering process is initiated where clusters are rebuilt, and new CHs are selected based on fresh Q values and energy measures. This provides network connectivity, load balancing, and energy fairness at the nodes, thus prolonging the network lifetime.

The incorporation of periodic reclustering with adaptive multihop routing makes the suggested approach resilient to topology variations in dynamic conditions, node failures, and energy imbalances typically found in underwater networks.

4.6 Termination criteria

When any one of the following criteria is met, the network operation is stopped: (i) the percentage of dead nodes exceeds a predetermined threshold, indicating significant node depletion; (ii) the total residual energy across the network drops below a minimum level set by the system; or (iii) the simulation reaches a maximum number of transmission rounds.

The Q-learning Bellman "(8)" was used to update the Q values. This equation states that each node modifies its Q value according to the projected future results and the immediate reward. The learning rate, α, was adjusted to 1 to fully emphasize the most recent reward. This ensures that the Q value updates are based only on the most recent observed reward, which is appropriate for dynamic underwater situations where past information may rapidly become irrelevant. The discount factor, γ, was set at 0.95, a widely accepted figure that emphasizes the present while retaining sensitivity to future benefits [19]. A neutral, untested state–action combination was represented by an initial Q value of 0.5, which is a common technique in Q-learning-based protocols. Additionally, a learned optimal path in the subsequent state was represented by max Q set to 0.9, which is consistent with the optimistic value initialization observed in the reinforcement learning literature [8, 10].

A node is considered "dead" when its energy falls to zero. Clusters are reconstructed based on updated Q values and energy conditions, and cluster heads are chosen from among nodes that are still alive. Dynamic clustering and reclustering are triggered when a cluster head's energy drops below a predetermined threshold value. The suggested routing technique for UWSNs based on reinforcement learning is described in Figure 2.

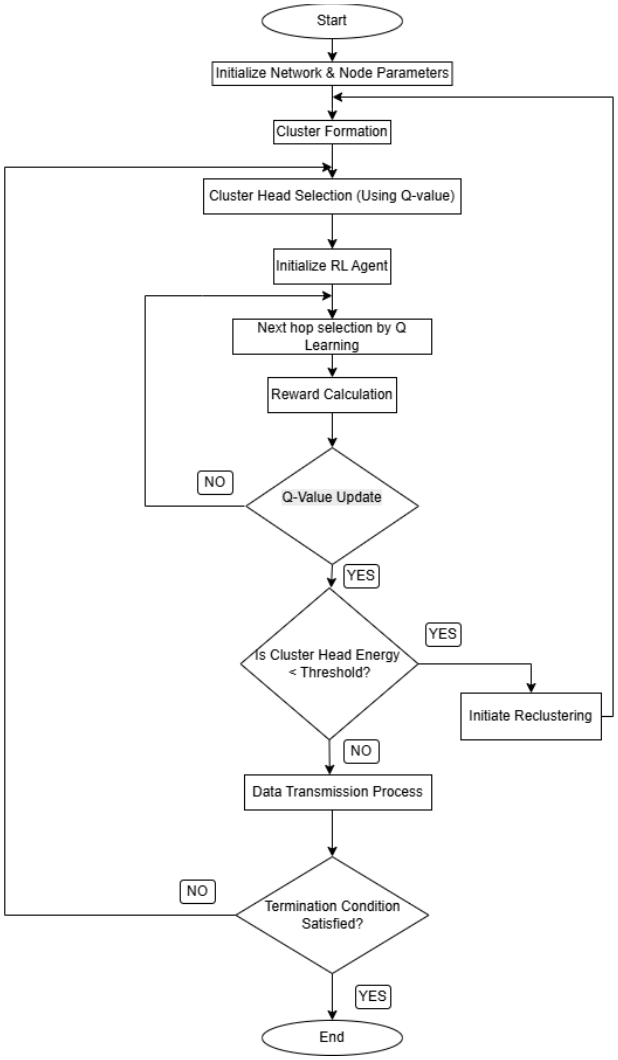

Fig. 2. Flowchart of the proposed reinforcement learning-based routing algorithm

I. RESULT AND DISCUSSION

To analyze the performance of the suggested reinforcement learning-based routing approach, MATLAB simulations were performed in an underwater environment of size 100 m × 100 m × 100 m [26]. There were 100 sensor nodes that were randomly scattered in the environment, including anchored nodes attached to the seabed and relay nodes hanging at different depths. One surface sink node was located at the top center of the network coordinates (50, 50) with both radio frequency and acoustic modules for communication with underwater sensor nodes and exterior monitoring stations.

The sensor nodes were initialized with the following: position coordinates (x, y, z), initial energy of 1 Joule, hop count to the sink node, communication range, sensing radius, and node status (active or dead). The model for energy consumption given in Section 4.1 was utilized, which covers energy for reception, transmission, and aggregation of data. The size of the data packet was taken as 4000 bits. The amount of energy for transmission or reception of a single bit was 50 nJ/bit, and the energy for data aggregation was 5 nJ/bit. The transmission amplifier energy was taken as 100 pJ/bit/m² based on the distance-squared propagation model [13,15]. The propagation velocity of the acoustic signals was assumed to be 1500 m/s [6]. During 5000 simulation rounds, nodes update their Q values dynamically and choose routes according to residual energy and environmental feedback, enabling adaptive and energy-efficient multihop communications.

The suggested RL-based method for sensor node routing topology is depicted in Figure 3. The sink node is indicated by the green star, selected cluster heads are indicated by red circles, and sensor nodes are represented by blue circles. The active communication linkages created by Q-value-based decision-making that consider transmission success, distance, and residual energy are shown by the blue lines. The efficiency of the reinforcement learning mechanism in dynamic underwater environments is validated via adaptive clustering and multihop routing methods, which yield dependable and energy-efficient data transfer.

Figure 4 shows the concentration of dead nodes with increasing number of rounds. If the energy of a node drops below a threshold, the node is marked as dead, and upon the death of a node, it should be excluded from the network or not allowed to participate in subsequent operations, thereby addressing energy management. It depicts the cumulative number of dead nodes over simulation rounds, highlighting the energy depletion trend in the network. Initially, no node dies until approximately the 2000th round, indicating balanced energy usage due to RL-based routing. As the number of rounds increases, the number of dead nodes increases steadily, with a sharp increase observed after 2500 rounds. This pattern reflects increased energy consumption due to prolonged operation and data forwarding. The figure demonstrates the network's sustained performance up to the middle of the simulation and validates the efficiency of the proposed method in delaying node failure and extending the network lifetime.

Over simulated rounds, the evolution of operational nodes is shown in figure 5. During the first phase, all nodes stay active, but when energy is depleted, there is a slow, stepwise decrease. 25 nodes are still active after 5000 rounds, meaning that 75 nodes have used up all their energy. Underwater sensor network environments frequently experience discrete node failure events, which are represented by the curve's stepwise shape.

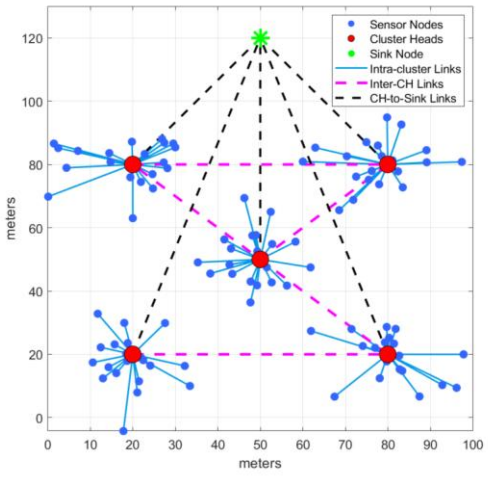

Fig. 3. Reinforcement learning-based routing topology for an underwater wireless sensor network, showing acoustic multihop communication between sensor nodes and selected cluster heads toward the surface sink.

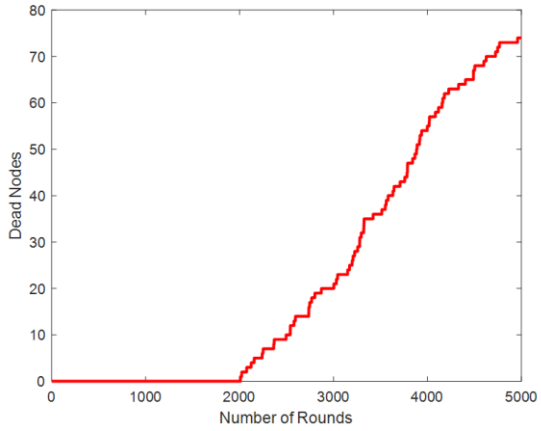

Fig. 4. Cumulative number of dead nodes as a function of simulation rounds.

The energy usage per round during the simulation is shown in figure 6. There is efficient energy utilization and balanced load distribution in the network when the dead nodes are minimal. As the number of rounds increases, the rate of node failure accelerates due to energy depletion caused by continuous data transmission and routing operations This figure shows how node failure over time directly affects network activity and verifies the effectiveness of the RL approach in controlling energy consumption.

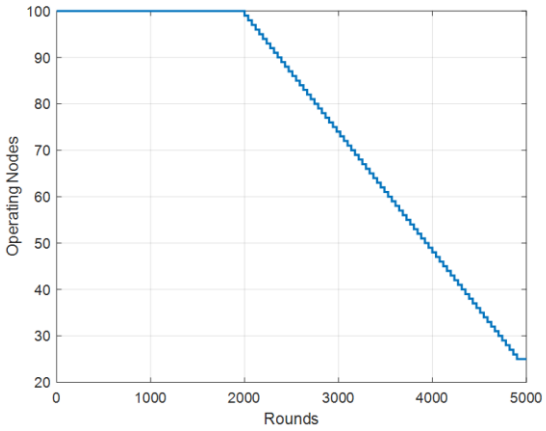

Fig. 5. Evolution of Operating Nodes over Simulation Rounds under Energy Depletion (100-Node UWSN).

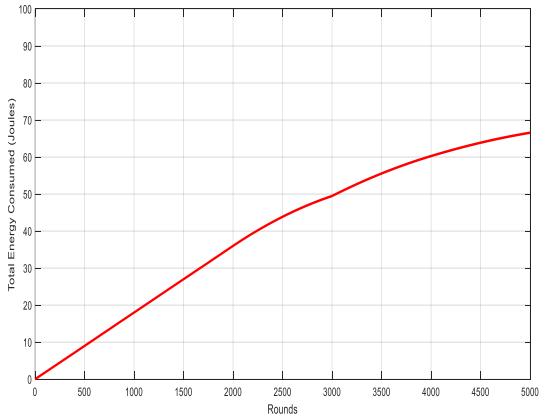

Fig. 6. Average network energy consumption per simulation round.

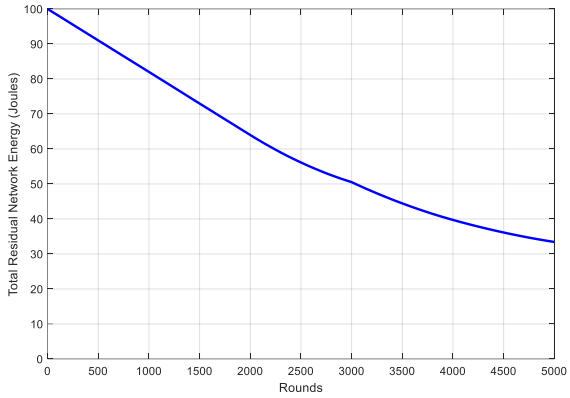

Fig. 7. Evolution of Total Residual Network Energy over Simulation Rounds under the Proposed Reinforcement Learning-Based Energy-Aware Routing Protocol in UWSNs.

The total residual network energy for the suggested RL-based routing protocol is depicted in the figure 7 during simulation rounds. The steady decline shows that energy-aware clustering and adaptive Q-learning-based routing were successful in achieving balanced energy use. The efficiency of the suggested approach in extending network lifetime under dynamic underwater settings is confirmed by the lack of sudden energy depletion.

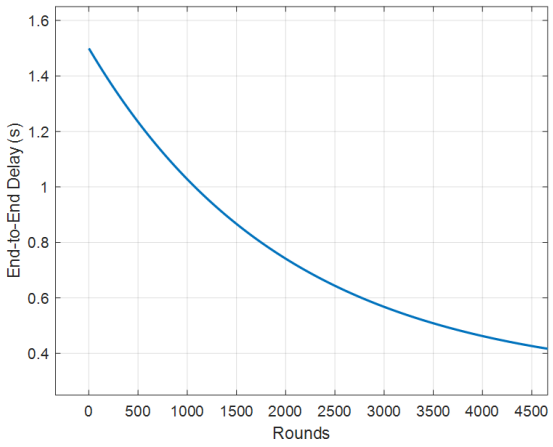

Fig. 8. Average End-to-End Communication Delay across Simulation Rounds under the Proposed RL-Based Routing Strategy in UWSNs.

The suggested reinforcement learning-based routing protocol's average end-to-end delay over simulation rounds is displayed in the figure 8. Because routing paths are not yet optimized during the Q-learning exploration phase, the initial rounds have a larger delay. Nodes choose dependable and energy-efficient multihop paths as learning advances, which steadily reduces delay. Despite node failures, the latency stabilizes at a lower value in subsequent rounds, demonstrating the learning process' convergence and the suggested routing scheme's resilience in dynamic underwater environments.

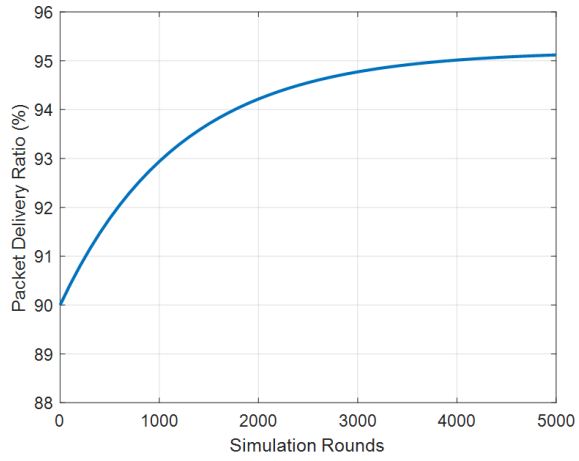

Fig. 9. Packet Delivery Ratio vs. Simulation Rounds for RL-Based UWSN Routing Protocol.

As simulation time increases, the Packet Delivery Ratio (PDR) steadily improves, although in the early rounds it rises quickly because of the system's quick adaptability. The improvement slows down and the PDR approaches a constant value near 95.2 % after about 2,000 rounds. This pattern shows that the system achieves a steady operating condition with reliable packet delivery over an extended period as shown in figure 9.

The efficiency of the suggested RL-based routing method was further confirmed by comparing it to two well-known protocols: RLOR and VBF [15]. Comparisons are performed on important performance metrics, such as average energy consumption, the PDR, network lifetime, and the number of dead nodes at the conclusion of the simulation.

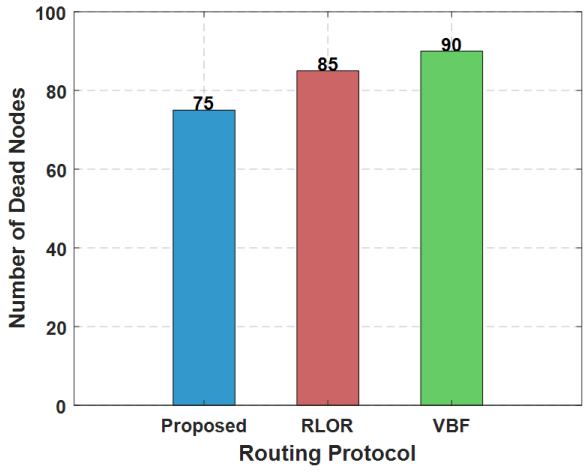

Fig. 10. Comparative Analysis of Dead Nodes under Different Routing Protocols after 5000 Simulation Rounds.

The number of dead nodes following 5000 simulation rounds for the suggested routing protocol, RLOR, and VBF are compared in Figure 10. Compared to RLOR (85) and VBF (90), the suggested approach displays fewer dead nodes (75), indicating enhanced energy balance attained through energy-aware clustering and adaptive Q-learning-based routing.

In the present research, the energy of data aggregation is approximated as a constant for all cluster heads, relying on the assumption that a uniform packet size of 4000 bits is used and that there exists a fixed per-bit aggregation cost of 5 nJ/bit—a practice commonly used in UWSN energy models [13,15]. Each cluster head is assumed to perform a single aggregation function per round, either its own sensed data or an aggregated packet from a small group of member nodes. This fixed-cost model offers an efficient first-order approximation that simplifies simulation complexity while providing fair and consistent routing performance evaluation.

Table I shows the energy consumption distributions for five sample sensor nodes that serve as cluster heads in a communication round. All nodes start with a capacity of 1 Joule. The transmission energy varies with the node-to-sink distance, as expected, because of the acoustic path loss model, whereas the reception and data aggregation energy does not vary across nodes. These values capture the reduced energy model specified in Section 4.1.2 and provide direct support for the RL-based routing methodology by demonstrating how energy usage and residual energy change with location and hence affect Q value updates and routing. Although aggregation energy might differ in actual field deployments according to cluster size and complexity of data processing, this difference is not regarded here as being within the scope of the present work and is a subject proposed for further improvement. The learning outcomes for each node are represented by the values in Table II, which are obtained from the RL mechanism explained in Section 4.4.2.

The reward of each node, which is portrayed as an autonomous agent, is determined by the distance to the sink, the success or failure of its most recent transmission, and its current residual energy. Energy-aware routing models for UWSNs typically assume that all nodes are initialized with a normalized energy value of 1 J to ease proportional reward computation [15]. Section 4.1's communication energy model was used to determine residual energy values.

A flag value of 1 indicates a successful delivery confirmed via an ACK, whereas a value of 0 indicates a failure because of packet loss, poor link quality, or delay. These two fields were assigned according to the packet delivery outcome. Likewise, if a transmission failed or was delayed, the penalty was set to 1; otherwise, it was set to 0. Both terrestrial and underwater wireless sensor networks have long used reinforcement learning-based routing strategies, which are compatible with the usage of binary success and penalty indicators [8, 13].

The weights used in the reward function, specifically $\boldsymbol{w_1}$ = 0.3, $\boldsymbol{w_2}$ = 0.2, $\boldsymbol{w_3}$, = 0.4, and $\boldsymbol{w_4}$ = 0.1, were chosen to offer a fair trade-off between transmission reliability, energy efficiency, and closeness to the sink. In line with the performance goals covered in [15,13,10], these values provide more weight to reliable packet delivery while still taking transmission distance and energy efficiency into consideration.

Table III presents an overview of the findings. The performance metrics for the RLOR and VBF protocols are obtained from relevant literature sources. These values are included for comparative purposes to highlight the relative improvements achieved by the proposed RL-based routing algorithm.

TABLE I
ENERGY CONSUMPTION AND RESIDUAL ENERGY COMPUTATION FOR SENSOR NODES ON THE BASIS OF THE NODE-TO-SINK DISTANCE

Node ID	Distance to Sink (m)	Transmission Energy (J)	Reception Energy (J)	Data Aggregation Energy (J)	Total Energy Used (J)	Residual Energy (J)
N1	25	0.00045	0.00020	0.00002	0.00067	0.99933
N2	30	0.00056	0.00020	0.00002	0.00078	0.99922
N3	45	0.00101	0.00020	0.00002	0.00123	0.99877
N4	50	0.00120	0.00020	0.00002	0.00142	0.99858
N5	60	0.00164	0.00020	0.00002	0.00186	0.99814

TABLE II
Q-VALUE UPDATE BASED ON THE REWARD FUNCTION AND RL PARAMETERS

Node ID	Distance to Sink (m)	Residual Energy (J)	Success_Flag	Penalty	Rewardr_t	Updated Q-Value
N1	25	0.99933	1	0	0.7078	1.5628
N2	30	0.99922	1	0	0.7065	1.5615
N3	45	0.99877	1	0	0.7041	1.5591
N4	50	0.99858	1	0	0.7036	1.5586
N5	60	0.99814	0	1	0.2028	1.0578

As evident from Table III, the proposed reinforcement learning-based routing protocol performs better than VBF and RLOR in terms of all performance aspects because of the collaborative application of energy-aware clustering and Localized Q-learning-based routing. A longer network lifetime of 5000 rounds is achieved as compared to VBF with 4200 rounds and RLOR with 4700 rounds, as energy consumption is balanced by the periodic re-clustering technique with residual energy-based relay node selection, thereby avoiding the earliest possible death of critical nodes. Additionally, the enhanced packet delivery ratio of 95.2% is achieved because of the adaptive rewards scheme that promotes reliable paths with a penalty on failed or energy-wastage transmissions in contrast to VBF's geometric routing and Opportunistic routing of RLOR. Finally, the lowest average energy consumption of 68 J is achieved by minimizing transmissions with large distances using the cluster-based aggregation technique along with next hop routing using reinforcement learning, ensuring that energy is effectively spent with only 75 nodes dying at the end of the simulation period.

TABLE III
PERFORMANCE COMPARISON BETWEEN THE PROPOSED METHOD AND EXISTING PROTOCOLS

Protocol	Network Lifetime (Rounds)	Packet Delivery Ratio (%)	Avg. Energy Consumption (J)	Dead Nodes at 5000 Rounds
Proposed (RL-Based)	5000	95.2	68	75
RLOR [14]	4700	91.8	75	85
VBF [31]	4200	88.5	82	90

CONCLUSION

This paper proposed a RL-based routing scheme for UWSN that combined energy-efficient clustering with locally optimized Q-learning-based routing to overcome the challenges of energy constraint, large propagation delay, and dynamic network topology. Simulation experiments show that the proposed scheme substantially outperforms existing schemes in improving network performance. Specifically, the proposed scheme achieved a network lifetime of 5000 rounds, which is an improvement of 19% over VBF and 6.4% over RLOR. It also achieved a packet delivery ratio of 95.2%, which is 7.6% higher than VBF and 3.4% higher than RLOR, indicating better routing integrity. Additionally, it reduced the average energy by 68 J, which is a lower energy consumption of 20.6% and 13.3% relative to VBF and RLOR, respectively, by adaptive energy-efficient clustering and optimized next hop routing learning. Thus, with more balanced energy consumption, the number of dead nodes at the end of the simulation is restricted to 75, which is lower than 85 for RLOR and 90 for VBF, ensuring better routing load distribution with delayed node death. These numerical experiments clearly attest that the proposed RL-based routing solution for reliable and energy-efficient underwater communication is effective. Future work would further improve the scheme with both deep reinforcement learning and multi-agent reinforcement learning for better scalability and flexibility in three-dimensional underwater networks.

ACKNOWLEDGMENTS

The authors thank the University of Mysore for the support provided during the research.

REFERENCES

[1] I. F. Akyildiz, D. Pompili, and T. Melodia, “Underwater acoustic sensor networks: Research challenges,” Ad Hoc Netw., vol. 3, no. 3, pp. 257–279, 2005. https://doi.org/10.1016/j.adhoc.2005.01.004

[2] I. F. Akyildiz, D. Pompili, and T. Melodia, “State-of-the-art in protocol research for underwater acoustic sensor networks,” in Proc. ACM Int. Workshop Underwater Networks (WuWNet), Los Angeles, CA, USA, 2006, pp. 7–16. https://doi.org/10.1145/1161039.1161043

[3] J. Partan, J. Kurose, and B. N. Levine, “A survey of practical issues in underwater networks,” in Proc. ACM Int. Workshop Underwater Networks (WuWNet), Los Angeles, CA, USA, 2006, pp. 17–24. https://doi.org/10.1145/1161039.1161045

[4] M. Ayaz and A. Abdullah, “Underwater wireless sensor networks: Routing issues and future challenges,” in Proc. 7th Int. Conf. Advances in Mobile Computing and Multimedia (MoMM), Kuala Lumpur, Malaysia, 2009, pp. 370–375. https://doi.org/10.1145/1821748.1821819

[5] W. Liu, et al., “Clustering-based reinforcement learning routing protocol with low complexity for underwater acoustic sensor networks,” in Proc. IEEE/CIC Int. Conf. Commun. China Workshops (ICCC Workshops), Xiamen, China, 2021, pp. 200–204. https://doi.org/10.1109/icccworkshops52231.2021.9538885

[6] H. Yu, N. Yao, and J. Liu, “An adaptive routing protocol in underwater sparse acoustic sensor networks,” Ad Hoc Netw., vol. 34, pp. 121–143, 2015. https://doi.org/10.1016/j.adhoc.2014.09.016

[7] X. Li, X. Hu, W. Li, and H. Hu, “Routing Protocol Design for Underwater Optical Wireless Sensor Networks: A Multiagent Reinforcement Learning Approach,” in Proc. IEEE Int. Conf. Commun. (ICC), Shanghai, China, 2019, pp. 1–7. https://doi.org/10.1109/jiot.2020.2989924

[8] S. Basagni, V. Di Valerio, P. Gjanci, and C. Petrioli, “MARLIN-Q: Multimodal communications for reliable and low-latency underwater data delivery,” Ad Hoc Netw., vol. 82, pp. 134–145, Aug. 2018. https://doi.org/10.1016/j.adhoc.2018.08.003

[9] S. Basagni, V. Di Valerio, P. Gjanci, and C. Petrioli, “Finding MARLIN: Exploiting multimodal communications for reliable and low-latency underwater networking,” in Proc. IEEE INFOCOM, Atlanta, GA, USA, 2017, pp. 1–9. https://doi.org/10.1109/infocom.2017.8057132

[10] Y. Chen, J. Bai, and Y. Li, “PDDQN-HHVBF routing protocol based on empirical priority DDQN to improve HHVBF,” Electronics, vol. 11, no. 24, p. 4031, Dec. 2022. https://doi.org/10.3390/electronics11234031

[11] A. F. E. Abadi, S. A. Asghari, M. B. Marvasti, G. Abaei, M. Nabavi, and Y. Savaria, “RLBEEP: Reinforcement-learning-based energy efficient control and routing protocol for wireless sensor networks,” IEEE Access, vol. 10, pp. 44123–44135, 2022. https://doi.org/10.1109/access.2022.3167058

[12] P. Kumar and C. Asbe, “An energy efficient routing algorithm for WSN using Q-learning based data aggregation method,” in Proc. IEEE 4th India Council Int. Subsections Conf. (INDISCON), 2023. https://doi.org/10.1109/indiscon58499.2023.10270449

[13] Y. Wang, Y. Li, J. Lei, and F. Shang, “Robust and energy-efficient RPL optimization algorithm with scalable deep reinforcement learning for IIoT,” Comput. Netw., vol. 255, p. 110894, Jan. 2024. https://doi.org/10.1016/j.comnet.2024.110894

[14] Y. Zhang, Z. Zhang, L. Chen, and X. Wang, “Reinforcement learning-based opportunistic routing protocol for underwater acoustic sensor networks,” IEEE Trans. Veh. Technol., vol. 70, no. 3, pp. 2756–2770, Mar. 2021. https://doi.org/10.1109/tvt.2021.3058282

[15] M. S. Hajar, H. K. Kalutarage, and M. O. Al-Kadri, “3R: A reliable multiagent reinforcement learning based routing protocol for wireless medical sensor networks,” Comput. Netw., vol. 237, p. 110073, 2023. https://doi.org/10.1016/j.comnet.2023.110073

[16] D. Prabhu, R. Alageswaran, and S. M. J. Amali, “Multiple agent-based reinforcement learning for energy efficient routing in WSN,” Wireless Netw., vol. 29, pp. 1787–1797, 2023. https://doi.org/10.1007/s11276-022-03198-0

[17] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018. https://doi.org/10.1007/978-1-4615-3618-5_1

[18] H. Yan, Z. J. Shi, and J. H. Cui, “DBR: Depth-based routing for underwater sensor networks,” in Proc. Int. Conf. NETWORKING Ad Hoc and Sensor Networks, Wireless Networks, Next Generation internet, Singapore, 2008, pp. 72–86. https://doi.org/10.1007/978-3-540-79549-0_7

[19] M. Ayaz, et al., “A survey on routing techniques in underwater wireless sensor networks,” J. Netw. Comput. Appl., vol. 34, pp. 1908-1927, Nov. 2011. https://doi.org/10.1016/j.jnca.2011.06.009

[20] Singh, S.B., Rizvi, M., Saxena, K. et al. “An adaptive, energy-efficient and secure routing protocol for zone-related mobile Ad-hoc networks using reinforcement learning”, Sci Rep (2025). https://doi.org/10.1038/s41598-025-32918-7

[21] Y. Zhang et al., “Multi-agent deep reinforcement learning based multiple access for underwater cognitive acoustic sensor networks,” computers and electrical engg, vol. 120, part C , p. 109819, Dec. 2024. https://doi.org/10.1016/j.compeleceng.2024.109819

[22] Sajid Ullah Khan, “Energy-efficient routing protocols for UWSNs: A comprehensive review of taxonomy, challenges, opportunities, future research directions, and machine learning perspectives”, Jrnl of King Saud University - Computer and Info Sci, Vol 36, Issue 7, 2024, 102128, https://doi.org/10.1016/j.jksuci.2024.102128

[23] Y Sun et al, “Adaptive clustering routing protocol for underwater sensor networks, Ad Hoc Networks, Vol 136, 2022, 102953, https://doi.org/10.1016/j.adhoc.2022.102953

[24] Li, Z., Chitre, M. & Stojanovic, M. Underwater acoustic communications. Nat Rev Electr Eng 2, 83–95 (2025). https://doi.org/10.1038/s44287-024-00122-w

[25] R Sharma, V Vashisht, U Singh “Modelling and simulation frameworks for wireless sensor networks: a comparative study,” IET, Oct. 2020. https://doi.org/10.1049/iet-wss.2020.0046

[26] R T Rodoshi et al., “Reinforcement Learning-Based Routing Protocol for Underwater Wireless Sensor Networks: A Comparative Survey,” IEEE Access, pp. 99, Nov 2021. https://doi.org/10.1109/ACCESS.2021.3128516

[27] R Singh and A Jain, “Deep Reinforcement Learning Enhanced Geographic and Cooperative Opportunistic Routing Protocol for Underwater Wireless Sensor Networks,” IJISAE., vol. 11, 2023. https://orcid.org/0000-0001-7424-7039

[28] Nguyen NT, Le TTT, Nguyen HH, Voznak M. “Energy-Efficient Clustering Multi-Hop Routing Protocol in a UWSN”. Sensors 2021 Jan 18;21(2):627. doi: 10.3390/s21020627.

[29] Z. Liu, Z. Sun, J. Su, Y. Yuan and X. Guan, "TBR: Secure Routing Design for UWSN Based on Trust Management Models," in IEEE Internet of Things Journal, vol. 12, no. 17, pp. 36674-36685, 1 Sept.1, 2025, https://doi.org/10.1109/JIOT.2025.3583776

[30] Cui et al., “Deep reinforcement learning-based adaptive modulation for OFDM underwater acoustic communication system” EURASIP Journal on Advances in Signal Processing (2023) 2023:1 https://doi.org/10.1186/s13634-022-00961-5

[31] P. Xie, J.-H. Cui, and L. Lao, “VBF: Vector-Based Forwarding Protocol for Underwater Sensor Networks,” in Proc. IFIP Networking, 2006.