

Dynamic Clustering Method for Underwater Wireless Sensor Networks based on Deep Reinforcement Learning

Kohyar Bolvary Zadeh Dashtestani¹, Reza Javidan¹ and Reza Akbari¹

Received: 20 November 2023 / Accepted: 14 October 2024
© Harbin Engineering University and Springer-Verlag GmbH Germany, part of Springer Nature 2025

Abstract

Underwater wireless sensor networks (UWSNs) have emerged as a new paradigm of real-time organized systems, which are utilized in a diverse array of scenarios to manage the underwater environment surrounding them. One of the major challenges that these systems confront is topology control via clustering, which reduces the overload of wireless communications within a network and ensures low energy consumption and good scalability. This study aimed to present a clustering technique in which the clustering process and cluster head (CH) selection are performed based on the Markov decision process and deep reinforcement learning (DRL). DRL algorithm selects the CH by maximizing the defined reward function. Subsequently, the sensed data are collected by the CHs and then sent to the autonomous underwater vehicles. In the final phase, the consumed energy by each sensor is calculated, and its residual energy is updated. Then, the autonomous underwater vehicle performs all clustering and CH selection operations. This procedure persists until the point of cessation when the sensor's power has been reduced to such an extent that no node can become a CH. Through analysis of the findings from this investigation and their comparison with alternative frameworks, the implementation of this method can be used to control the cluster size and the number of CHs, which ultimately augments the energy usage of nodes and prolongs the lifespan of the network. Our simulation results illustrate that the suggested methodology surpasses the conventional low-energy adaptive clustering hierarchy, the distance- and energy-constrained K-means clustering scheme, and the vector-based forward protocol and is viable for deployment in an actual operational environment.

Keywords Underwater wireless sensor network; Clustering; Cluster head selection; Deep reinforcement learning

1 Introduction

Recent evolutions in wireless communication-based systems have led to a massive expansion in the field of under-

water wireless sensor networks (UWSNs). Reliable infrastructures must be established to manage and plan these networks for sustainable communication among the computer world, the undersea world, and the surrounding terrestrial environment (Akyildiz et al., 2002; Nguyen et al., 2021). UWSNs involve the random deployment of a large number of autonomous sensing nodes with limited battery energy, which are considered homogeneous in the depth of underwater area (Fattah et al., 2020; Nguyen et al., 2021). Underwater acoustic communication is ideal for long-distance data transmission because of the harsh underwater conditions; thus, a high transmission power is required to convey the obtained data between sensors (Rowshanrad et al., 2014; Stojanovic and Preisig, 2009).

To this end, certain issues are considered in the design of these networks to decrease power consumption to increase the network lifetime. The findings of numerous research show a coherent and programmed sensor grouping can considerably aid in resolving the scale issue in sensor networks. In this method, each group of sensors is referred to as a cluster, and each cluster contains a node that is selected as the precursor and referred to as a cluster head (CH) (Rowshanrad et al., 2014). CHs are elected to organize data gath-

Article Highlights

- The study introduces a dynamic clustering method for 3D-UWSNs using deep reinforcement learning and partially observed Markovian decision process models to enhance energy efficiency.
- The integration of the elbow method with k-means clustering ensures optimal cluster sizes, facilitates efficient data transmission and minimizes communication overhead.
- A multistep DRL-based approach is proposed for selecting cluster heads, maximizing energy conservation by considering node energy levels and proximity to cluster centroids.
- Simulation results demonstrate that the proposed method achieves superior energy efficiency and prolongs network lifetime, validating its practical applicability in real-world underwater environments through effective data collection by autonomous underwater vehicles.

✉ Reza Javidan
javidan@sutech.ac.ir

¹ Department of Computer Engineering and Information Technology, Shiraz University of Technology, Shiraz 7155713876, Iran

ering from sensors located within their respective clusters. When gathering information from each underwater sensor node, they employ multi-access mechanisms in a distributed or centralized manner to prevent clashes. CH selection is centralized by a base station (BS), but in the distributed version, the process is self-organized by each sensor. In addition, CHs combine the gathered data by eliminating duplicates and sending them to the BSs. Energy loss is guaranteed through the frequent rotation of CHs (Bholowalia and Kumar, 2014; Luo et al., 2021; Taheri et al., 2016).

Traditional methods for clustering UWSNs involve consensus among sensing nodes to determine the CH, which then aggregates and transmits data to the next CH. However, this process is plagued by considerable communication overheads during the selection and maintenance of the CH, particularly in dynamic networks, and the energy consumption required for long-distance communication by the CH (Khan et al., 2018; Majid et al., 2016). On the contrary, the novel clustering techniques utilized integrate sophisticated methodologies, such as machine learning algorithms, meta-heuristics, and reinforcement learning (RL), to augment cluster establishment, routing effectiveness, and energy preservation. These emerging methodologies exploit pioneering strategies to address issues, including coverage gaps, sensor node mobility, and environmental variations, which offer flexible and effective resolutions for underwater communications. In essence, although old-fashioned clustering techniques establish a fundamental framework for network structuring, modern clustering methodologies present improved functionalities and flexibility to cater to the progressing requirements of underwater sensor networks (Gupta and Singh, 2024).

Data collection by autonomous underwater vehicles (AUVs) dramatically increases the lifespan and reduces the energy consumption of each sensor in such clustering approaches by providing assistance in data collection, storage, and computing services due to the shorter distances traveled. AUVs reduce the hotspot problem produced by several sensor nodes relaying their data to the same intermediate node (Jawhar et al., 2018; Majid et al., 2016).

This article presents a new approach for energy-efficient dynamic clustering in three-dimensional (3D)-UWSNs through the utilization of partially observed Markovian decision process (PO-MDP) models and deep RL (DRL) algorithms. The aim was to create scalable clusters with minimal power consumption, where CHs are selected in accordance with predetermined criteria. The Markov conditions must be satisfied, where the states at time $t+1$ and $t-1$ are exclusively dependent on the present state of the network at time t . Once these states have been defined, appropriate actions, such as CH selection and reclustering/not reclustering based on energy saving and proximity to the cluster's center, can be determined. The reward function of DRL was employed to evaluate the effectiveness of each

action, and a method that converges in fewer states through the selection of positive actions with higher rewards, which results in accurate CH election, was successfully proposed.

This paper contributes the following:

- Clustering of the 3D-UWSN using the k-means algorithm and elbow method to ensure the maximum number of clusters;
- Introduction of a new multistep energy-efficient and life span-aware DRL-based CH election method to alleviate the need for intense deployment of low energy-consuming sensors in UWSNs and overcome the hotspot problem;
- Enhancement of the energy efficiency of the entire network by the reward function, which considers the energy loss incurred by the CH during data transmission to the AUV, the remaining energy of the present CH, energy usage of non-CH sensors during data transmission to the CH, and the distance between the CH and the cluster's centroid.
- Utilization of various important metrics, including the energy levels of nodes and the proximity to the cluster centroid, to determine the CHs. In addition, the most favorable route for packet transmission to the AUV is identified to achieve reduced energy consumption, extended network lifespan, and enhanced network scalability.

In this study, experiments were conducted to evaluate and compare the proposed routing method with existing ones from the viewpoint of energy efficiency and network lifetime via simulations, which all demonstrated the effectiveness of the proposed method.

This paper is structured in the following manner. Section 2 presents the related works, and Section 3 delves into the system model and the implementation of the proposed model. Section 4 evaluates the proposed method and compares it with other methods. The last section presents the conclusions and suggestions that can lead to further improvements in the performance of sensor networks.

2 Related work

Various efficient routing protocols have emerged for UWSNs. This article briefly presents the key features, benefits, and drawbacks of these protocols. UWSN routing protocols can be categorized into three distinct classifications: energy-, database-, and geographic information-oriented protocols. Our proposed technique is focused on energy-oriented protocols, whose brief overview will be provided in this section (Khalid et al., 2017).

Energy-oriented protocols comprise five categories, including bio-inspired, cooperative-reliability-based, RL-based, depth-based, and cluster-based protocols. As our work involves clustering methods and RL-based routing protocols, we briefly introduce these protocols (Stojanovic and Preisig, 2009).

2.1 RL-based routing protocols

RL-based routing mechanisms for UWSNs are guided by engagements between humans and their environment. RL is focused on the actions of intelligent agents in intricate environments, where they engage with the environment to accomplish specific goals. This strategy involves comprehending the environment and defining reward criteria. In opposition to predetermined actions, RL agents learn through trial and error and adjust their behavior based on feedback from the environment. This adaptable characteristic enables RL-based routing protocols to optimize routing choices in changing underwater conditions to improve network performance and efficiency (Gupta and Singh, 2024; Wang et al., 2023).

A study introduced a Q-learning-based energy-efficient and lifespan-aware routing protocol called QELAR (Hu and Fei, 2010). The Q-value in QELAR is determined through successful packet transmission. QELAR's reward function is identified through the evaluation of the remaining energy of sensors and power allocation. The protocol selects routes that exhibit a higher degree of residual power than the shortest pathway. Nevertheless, superfluous energy depletion may ensue due to packet overhearing despite an increase in network lifetime compared with the vector-based forwarding (VBF) protocol presented in another work (Javaid et al., 2017).

The energy-efficient depth-based opportunistic routing algorithm with the Q-learning EDORQ algorithm (Lu et al., 2020) uses Q-learning for routing. This algorithm comprises two primary components: the selection of initial candidates and their coordination. The selection phase has greedy and void recovery modes. Greedy mode selects forwarding candidates, and void recovery mode is for network voids. The best benefit is attained by selecting the closest sensor to the sink with the most power. Sensors obtain holding time based on Q-values. EDORQ outperforms other algorithms in terms of power reliability, delivery rate, and network overhead.

Alsalmán and Alotaibi (2021) proposed a balanced routing protocol established in machine learning (BRP-ML) for UWSNs; it incorporates RL (Q-learning) to manage data routing from decentralized sensors to the sink node while considering constraints, such as power restrictions, latency, and the void area predicament. The convention is organized around four distinct stages: initialization, discovery, clustering, and data forwarding. Simulations were conducted to illustrate that BRP-ML achieved a reduction in latency and an enhancement in energy efficiency compared with established routing protocols, such as QELAR and QL-EDR. The outcomes indicated that BRP-ML balances energy usage and delivery latency while tackling the void area challenge, which enhances the overall performance of UWSNs.

Sun et al. (2022) proposed an adaptive clustering routing protocol for UWSNs based on multi-agent RL (MLAR). The protocol aims to address the hotspot problem caused by the frequent use of dominant nodes, which leads to premature energy depletion and reduced network lifetime. By integrating RL and clustering algorithms, the proposed MLAR protocol optimizes routing decisions based on energy and location advantages while selecting relays and CHs. Simulation results demonstrated that MLAR outperformed traditional clustering algorithms and RL-based protocols in terms of energy efficiency, network lifetime, packet delivery rate, latency, and total energy consumption in UWSNs.

2.2 Cluster-based routing protocols

These protocols employ cluster segmentation with CHs to dynamically discover intercluster paths and reduce flooding traffic during route discovery. Furthermore, these methodologies aid in alleviating challenges, including void holes, topology changes induced by mobility, and fluctuations in the environment, which lead to enhanced stability and resilience of underwater communications (Gupta and Singh, 2024; Khisa and Moh, 2021).

One research introduced a clustering approach called energy-efficient routing for UWSNs-A Clustering Approach (EERU-CA) (Khan et al., 2015), and it involves the use of a specialized sensor that functions solely as the CH. The authors asserted that this method enhances the system's energy efficiency by ensuring the placement of the CHs at the most concise route to reach their respective cluster members; however, the practical deployment of such nodes may be challenging.

Souiki et al. (2015) proposed two novel energy-efficient routing protocols, namely, single-hop fuzzy-based energy-efficient routing (SH-FEER) and multihop fuzzy-based energy-efficient routing (MH-FEER), that rely on fuzzy C-means clustering to establish clusters and select CHs. However, SH-FEER transfers data straight to the sink, whereas the other method utilizes a multihop pathway. In addition, the authors conducted simulations on static and dynamic node topologies. The aforementioned protocols are unsuitable for networks with high density owing to the utilization of the clustering approach. Furthermore, the authors investigated several node topologies that are both static and dynamic.

Wang et al. (2015) proposed energy-efficient grid routing based on a 3D cube (EGRC) algorithm, which operates in a fully controlled area divided into cube-shaped clusters. The BS broadcasts a message to notify the nodes of the grid area. The nodes, in turn, transmit messages to their neighbors and inform them of their remaining power and directions from the sink and grid area. In a multihop procedure, the selection of the CH for data transmission is based on the sensor with the highest remaining energy and closest proximity to the sink. However, the selection of the

succeeding pioneer sensor is more complex and requires the calculation of a weight based on the remaining power and distance to the sink.

The 3D-UWSN (Das and Ameer, 2017) system uses a clustering technique based on a cube grid, where each minor grid acts as an autonomous cluster. The communication process encompasses the creation of optimized clusters and sequential data transmission in rounds. CH election was performed on a round-by-round basis, with the CH serving as the sensor with the highest residual energy. The optimal cluster determination is contingent upon the calculation of the sphere's volume, node count, and the intervening space separating the nodes and the sink. Single-hop communication enables data transmission from non-CH sensors to the CH within the same cluster, and multihop transmission proceeds from the CH to the BS.

Mazinani et al. (2018) presented a novel routing algorithm that enhances the VBF protocol through the dynamic adjustment of the routing pipe's radius based on environmental factors and node energy levels. This adaptive algorithm prioritizes energy-efficient paths and selects forwarding nodes based on proximity to the destination and energy levels. This algorithm demonstrated superior performance in energy efficiency, packet delivery rates, and adaptability through simulations in comparison with existing protocols; this condition ensures improved communication reliability in UWSNs.

Zhu and Wei (2018) proposed an energy-efficient routing protocol based on layers and unequal clusters (EERBLC) for UWSNs. This protocol leverages layered and unequal clusters to maximize energy savings. Specifically, the CH selection process relies on the calculation of waiting time, which is determined by considering the present power of the sensors. Moreover, the protocol ensures load balancing among nodes by updating the CH accordingly. However, the use of flooding techniques may lead to network congestion and substantial overhead in routing, which should be considered.

Khan et al. (2019) introduced a routing protocol called multilayer cluster-based energy efficiency (MLCEE) for UWSNs. The protocol's fundamental assumption is that sinks have an infinite energy supply, and sensors are arranged in several tiers on the ocean floor. The holding time for each node is determined by considering the remaining and initial power of each sensor. If a sensor has a short holding time, then it contains a substantial amount of remaining energy and is likely to become a CH soon. The Bayesian spam filtering method handles the CH election. For the resolution of the hotspot problem, nodes in the initial layer are permitted to send data directly to the sink. However, CH updates are disregarded.

In 2019, a cluster-based UWSN (CUWSN) technique (Bhattacharjya et al., 2019) incorporating a grid-based approach was introduced. This technique involves the

selection of CHs on their residual energy, with the nodes possessing higher levels of remaining power being designated as CHs. In addition, a coordinator sensor was implemented to facilitate internal communication between sensors and data transformation to the sink. However, the likelihood of premature failure of CH and coordinator persists despite the potential for an enhanced throughput that the CUWSN may offer.

Omeke et al. (2021) discussed the challenge in energy consumption of UWSNs and proposed a new clustering algorithm called distance- and energy-constrained K-means clustering scheme (DEKCS) to prolong the network's lifetime. The algorithm selects potential CHs based on their position in the cluster and residual battery level and dynamically updates energy thresholds to prevent network disconnection.

A proposed hybrid UWSN network architecture, known as ACOP-UWSN (Shelar et al., 2022), utilizes clustering of optical sensors only based on OSNR and residual energy criteria for CH selection, which results in a better performance than those of state-of-the-art methods through the use of an OSNR-based energy-efficient clustering algorithm.

Nazareth and Chandavarkar (2024) presented a new underwater routing protocol known as cluster-based multi-attribute routing (CMAR) to solve challenges in UWSNs, particularly in handling void nodes during routing. CMAR is a sender-based opportunistic routing protocol that utilizes the technique for order of preference by similarity to ideal solution (TOPSIS) to evaluate neighboring nodes based on various attributes. Unlike other protocols, CMAR clusters only a specified threshold number of nodes to reduce complexity. CMAR's performance was compared with the HydroCast protocol through MATLAB simulations in terms of node selection, cluster formation, void-node handling, and transmission reliability. Simulations revealed that the protocol aims to enhance transmission reliability and reduce the number of formed clusters.

Table 1 provides a brief analysis of two energy-conserving routing methodologies employed in UWSNs that are proximate to our proposed approach.

In this paper, we developed an energy-efficient and life span-aware clustering protocol based on Q-learning. Our method utilizes a reward function that considers each sensor's remaining energy and distance from the cluster centroid. Through the DRL process, the protocol selects CHs with higher residual energy and concise proximity to the center of the cluster compared with non-CH sensors.

3 System model and the proposed method

In this study, we randomly scattered nonstationary sensors across a 3D underwater network. The elbow method and k-means algorithm were utilized to group sensors into

Table 1 UWSN energy-aware routing protocols comprehension (RL-based, cluster-based, and the proposed method)

Protocol type	Protocol name	Year	Key Idea of the protocol	Results
Energy-oriented RL-based Protocols	EDORQ (Lu et al., 2020)	2020	The EDORQ process consists of two distinct phases—candidate selection and coordination—which enable the identification of void recovery and forwarding nodes	Less power consumption by each node, improved network longevity, and less end-to-end delay
	BRP-ML (Als Salman and Alotaibi, 2021)	2021	A balanced routing protocol is proposed based on machine learning that aims to address the challenges faced by UWSNs, such as energy consumption, routing issues, limited bandwidth, high error rates, and node localization	Extends the network lifetime, reduces energy consumption, balances energy distribution, and improves data delivery rates
	MLAR (Sun et al., 2022)	2022	An adaptive lifespan aware clustering routing protocol for UWSNs using multi-agent reinforcement learning. The protocol targets the hotspot issue in underwater sensor networks by effectively choosing cluster heads	Maximizes the network lifetime and routing efficiency and lowers energy consumption
Energy-oriented cluster-based Protocols	EERU-CA (Khan et al., 2015)	2015	The method proposed the use of a customized sensor that individually functions as the CH	Energy efficiency enhancement
	SH-FEER, MH-FEER (Souiki et al., 2015)	2015	Both protocols employ Fuzzy C-means clustering and CH selection, but while the latter directly transfers data to the sink node, the former uses a multihop route	Successful simulations on both static and dynamic node topologies
	EGRC (Wang et al., 2015)	2016	The method selects a CH for data transmission in a multihop manner based on remaining power and directions to the sink	Upholds the reliability of data transmission
	3D-UWSN (Das and Ameer, 2017)	2017	Utilizes a clustering approach based on a cube grid. The selection of CH is based on remaining energy, with single-hop data transmission	Optimized energy efficiency and network longevity
	Enhanced VBF (Mazinani et al., 2018)	2018	Attending an enhanced version of the VBF algorithm by addressing energy consumption, optimizing packet delivery and dynamically adjusting the routing pipe radius based on network conditions	Outperforms existing routing protocols in terms of energy consumption and packet delivery
	EERBLC (Zhu and Wei, 2018)	2018	The CH is modified in order to distribute the workload evenly between the sensors	Energy efficiency, congestion and heavy routing overhead
	MLCEE (Khan et al., 2019)	2019	The protocol allows the first layer nodes to directly send data to the sink, assuming unlimited power supply at the sink nodes and layered arrangement of nodes on the seabed	Great enhancement in network lifetime, energy consumption and data transmission amount
	CUWSN (Bhattacharjya et al., 2019)	2019	A grid-cluster-based approach with CH election on residual energy with a coordinator node being responsible for communications between nodes inside the cluster and data transfer to the sink	Energy efficiency
DEKCS (Omeke et al., 2021)	2021	The protocol suggests a clustering algorithm to select CHs based on their position and remaining battery level and adjusts energy thresholds to avoid network disconnection	Prolongs network lifetime, outperforms LEACH protocol and an optimized version of LEACH k-means	
	ACOP-UWSN (Shelar et al., 2022)	2022	CH election is based on OSNR and remaining power criteria	Residual energy efficiency
	CMAR (Nazareth and Chandavarkar, 2024)	2024	Proposing a sender-based, opportunistic routing protocol that uses the TOPSIS technique to evaluate neighboring nodes and form clusters based on specified thresholds	Improves transmission reliability, reduces the number of clusters, and minimizes the void nodes in routing
Proposed method	A Dynamic Clustering Protocol based on Deep Reinforcement Learning		A dynamic multistep energy-efficient and life span aware clustering protocol that conducts the optimal clustering using a combination of elbow method and k-means algorithm and performs CH election based on DRL techniques	Optimization of energy consumption and extension of network lifespan

clusters, which in turn aided in data reporting through the selection of CHs from among themselves. (Bholowalia and Kumar, 2014). We employed DRL to further improve the election process of CHs. This approach considers the energy levels of nodes and their positions within their respective clusters. Data aggregation was then performed between non-CH nodes and the precursor CH node inside each cluster. Finally, the collected data were retrieved by the AUV, which hovered in a depth coordinate of the underwater network to provide a wide coverage over CHs within their clusters. The generality of the proposed method is depicted in Figure 1.

3.1 Clustering method using k-means and the elbow method

The set of sensor nodes in our UWSNs, denoted as $X = \{x_1, x_2, x_3, \dots, x_n\}$, consisted of d-dimensional vectors. In addition, K-means clustering partitions N underwater sensors into k clusters ($k \leq N$).

Let $C = \{c_1, c_2, c_3, \dots, c_k\}$ be the set of clusters where each c_k refers to a d-dimensional vector representing sensors in the corresponding cluster, and j represents the number of sensors in each cluster and $j < N$.

The geometric center of the sensors within each cluster, denoted as $V = \{v_1, v_2, v_3, \dots, v_k\}$, is determined by the following (Wang et al., 2020):

$$\mu_j = \left(\frac{1}{N} \sum_{i=1}^N x_i \right) \tag{1}$$

The distance between a cluster sensor and the center of the cluster was calculated using the Euclidean distance formula:

$$D(x) = \sqrt{\left\| \sum_{i=1}^j (x_i - \mu_j) \right\|^2} \tag{2}$$

To begin with the elbow method, we initialized the esti-

imated of k_{opt} with specific values in the range of $[k_{min}, k_{max}]$, with $\left[1, \left[\sqrt{N} \right] + 1 \right]$. Through the utilization of the elbow method for the abovementioned range of k , the within-cluster sum of squares error (WSSE) was calculated to find the optimal point, where the rate of decrease in WSSE slows down considerably (Bholowalia and Kumar, 2014). The number of clusters at this step was considered the optimal number. The formula for calculating WSSE is as follows:

$$WSSE = \sum_{i=1}^k \sum_{x_i \in C_i} \|x_i - \mu_i\|^2 \tag{3}$$

Then, a brief pseudocode of the proposed methodology was carried out in Algorithms 1 and 2. K-means algorithm convergence was achieved when the cluster centers did not change significantly between k-means iterations or after a specific number of iterations. The following algorithm accounts for the limitations of the underwater environment as the transmission range presents crucial restrictions.

3.2 Proposed CH election method

In this paper, we used DRL to select optimal CHs in UWSNs based on energy efficiency (Kozlowski et al., 2018; Li, 2017). DRL problems can be formulated as a PO-MDP when the observation of the environment satisfies the Markov property. A PO-MDP is defined by a 6-tuple $(\mathcal{S}, \mathcal{A}, t, \mathcal{P}_a, \mathcal{R}_a, \gamma)$ (Nasir and Guo, 2019; Rahmani et al., 2022). To provide a more comprehensive explanation of the intricate mechanisms underlying the operation of DRL in the context of CH election within our UWSN, we described these six parameters in a more elaborate manner.

\mathcal{S} defines the state space that represents the current situation of the UWSN that includes two variables: sensor locations and remaining energy levels that change through time. For our approach, let the state space be $\mathcal{S} = \{s_1, s_2, s_3, \dots, s_{2^a}\}$. Here, we determined the decision time points (t) along with the transition probability:

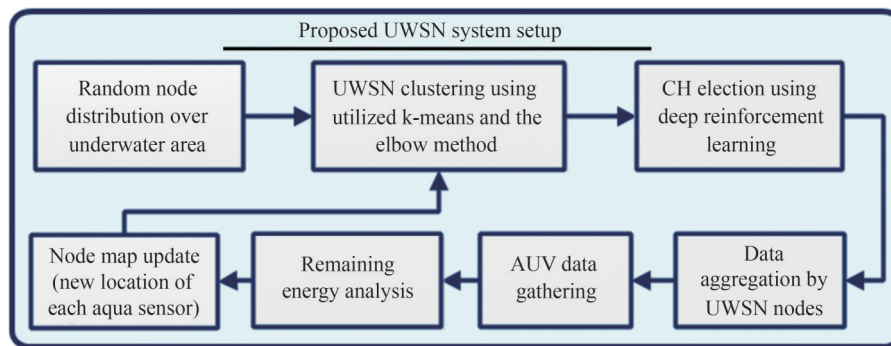


Figure 1 Implementation steps of our proposed algorithm

Algorithm 1 The elbow method in UWSN

Input: $X = \{x_1, x_2, x_3, \dots, x_n\}$ //set of N sensors
Output: k //optimal number of clusters

Create arrays WSS and k

for $k: = 1$ to $\lceil \sqrt{N} \rceil + 1$

begin

if ($k = 1$) then

$v_1(1): =$ centroid of X

$V_1: = \{v_1(1)\}$

else

for $i: = 1$ to N

Begin

Execute k-means ();

Calculate WSSE = $\sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2$

end

end if

$v_k(i): = v_k^i$, where, $1 \leq k \leq N$ and $WSSE_k = \min_{i=1}^N WSS_i$

end

Algorithm 2 The K-means algorithm in UWSN

Input: $X = \{x_1, x_2, x_3, \dots, x_n\}$ //set of N nodes
 K //Maximum number of clusters
Output: A set of k cluster

Define k-means ():

{

assign k as the initial cluster: //Initialization

$C = \{c_1, c_2, c_3, \dots, c_k\}$

while (convergence criteria=false)

{

assign nodes to the nearest cluster center based on the Euclidean distance $D(x)$ //Clusters update

$v_k^i = k^{\text{th}}$ cluster centroid obtained using //Centroid update

$\mu_j = \left(\frac{1}{N} \sum_{i=1}^N x_i, \frac{1}{N} \sum_{i=1}^N y_i \right)$

}

The final clusters and their respective cluster centers are outputted.

}

$$\mathcal{P}_{a, \mathcal{S}_t, \mathcal{S}_{t+1}}: \mathcal{S}_t \times \mathcal{A} \times \mathcal{S}_{t+1} \rightarrow [0, 1] \tag{4}$$

The equation above determines the likelihood that taking \mathcal{A} , which defines the action space in state \mathcal{S}_t , will result in state \mathcal{S}_{t+1} . \mathcal{A} represents the CH election actions that the agent can take. The set of actions should encompass all potential nodes that can be selected as the CH. This phase

involves the selection of a new CH with a higher energy than the non-CH nodes that are located at the closest proximity to the cluster’s center. Furthermore, the expected immediate reward, which is received at time t after transitioning from state \mathcal{S}_t to state \mathcal{S}_{t+1} , is factored into the following:

$$\left(\mathcal{P}_{\text{action}, \mathcal{S}_t, \mathcal{S}_{t+1}}: \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{R} \right) \tag{5}$$

where \mathcal{R}_a defines the reward function, such as minimizing energy consumption or maximizing data transmission rate, that should be designed to encourage the selection of a CH that maximizes network performance.

The aforementioned equation considers two-time steps and, subsequently, by applying action a_t and transitions from \mathcal{S}_t to \mathcal{S}_{t+1} , it assesses the reward based on the minimized difference between energy functions for these two states to promote the selection of a CH that maximizes network performance. The energy function is calculated using the following formula:

$$\left(E_{\text{energy-function}}(t) = \sum_{i=1}^N E_i(t) \right) \tag{6}$$

The discount factor (γ) ranging $[0, 1]$ is included to account for the value of future rewards. The graphical overview of the ever-changing interdependence between the agent (AUV) and its environment is presented in Figure 2.

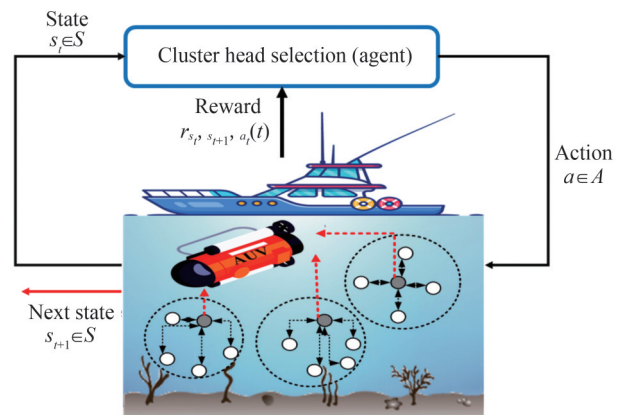


Figure 2 Interaction between the agent and the underwater area

The DRL agent mainly aims to find an optimal policy, $(\pi^*(s))$ that maximizes the discounted reward over time by implementing an appropriate action based on the current state. The agent is consistently furnished by the environment with the initial state and rewarded at the culmination of each iteration subsequent to the agent’s fulfillment of the requested action. This objective can be attained through the practical solution of the Bellman equation (Sutton and Barto, 2018):

$$Q_{\pi}^*(s, a) = \mathbb{E} \left[r(s, a) + \gamma \sum_{a'} \pi(a'|s) Q_{\pi}(s', a') \right] \quad (7)$$

According to the above equation, each state’s value is computed for the optimal action in every episode. Q-learning is an important advancement in RL that applies the Q-table approach and the Bellman equation:

$$Q(S_t, a_t) \leftarrow Q(S_t, a_t) + \alpha \left[r_t + \gamma \max_a \left(Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right) \right] \quad (8)$$

The Q-table trains the future reward associated with each action taken in a particular state and subsequently selects the action that offers the highest expected reward. Updating Equation 9 obtains the value function for each state action. The graphical overview of the Q-table is displayed in Figure 3.

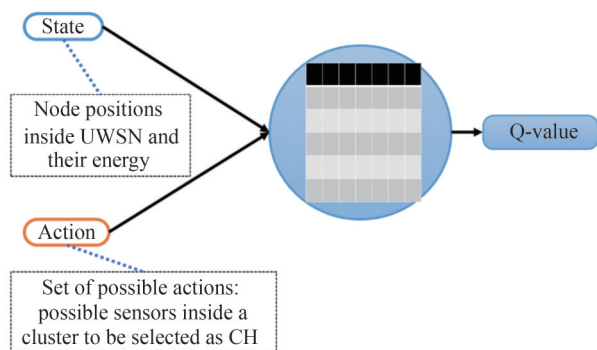


Figure 3 Graphical overview of the Q-table

Deep Q-learning (DQL) uses a deep neural network (DNN) to estimate nonlinear Q-values for every state action, which allows for an efficient approach. The agent establishes the optimal action policy by considering all possible actions in the environment and utilizes an epsilon-greedy algorithm to strike a balance, exploiting known information and exploring new possibilities (Balhara et al., 2022).

$$\pi^*(s) = \arg \max_{a \in A} \left[Q(s, a; \mathbf{w}^*) \right] \quad (9)$$

Experiences are stored in a replay buffer, and its weight vector (\mathbf{w}) is updated the agent by selecting a minibatch of occurrences from the buffer. The training process entails the adjustment of the neural network’s weights to reduce the disparity between the predicted Q-values and the real rewards using mean squared error. This step continues until the convergence of the model to a stable set of weights that accurately predicts the optimal actions for each state. The graphical overview of the DQL is illustrated in Figure 4.

After training, the DQL model can be used to select the optimal CH for a given state. This step involves feeding the current state of the network into the model and select-

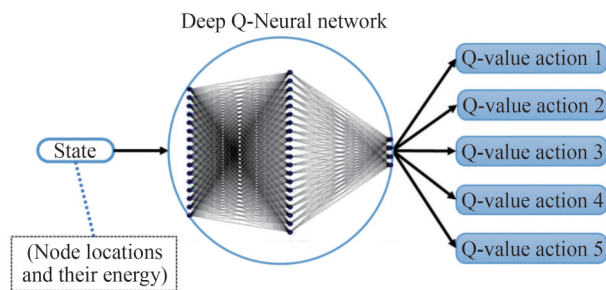


Figure 4 Graphical overview of the DQL

ing the action with the highest predicted Q-value. The selected node serves as the CH for that round, and the process is repeated in each subsequent round. The selection of nodes with high energy levels that are suited to become CHs can improve network performance and prolong the lifetime of the network. The DQL algorithm implemented in our proposed UWSN comprises a series of steps:

- a) Initialize the Q-network with random weights.
- b) Observe the current state (s) of the UWSN, including the number of active sensors, environmental locations, and the remaining energy that affects the network performance.
- c) Choose an action a based on an epsilon-greedy policy.
- d) Execute action a and observe reward r_a and the next state (s_{t+1}) based on the minimized energy consumption.
- e) Store the experience (s, a, r_a, s_{t+1}) in the replay buffer.
- f) Sample a batch of experiences from the replay buffer.
- g) Compute the target Q-values using the Bellman equation.
- h) Update the Q-network weights to minimize the mean squared error between the predicted and target Q-values.
- i) Repeat steps 2–8 until convergence.

Refer to (Arulkumaran et al., 2017; Balhara et al., 2022) for a detailed examination of DRL and its real-world applications.

As previously mentioned, energy consumption occurs, and sensors collect, analyze, and transmit data. The present study highlighted that power consumption during packet transmission depends on the packet size and range of transmission.

For sending data with k -bit size over a distance d , the necessary power required by the transmitter $E_{trans}(k, d)$ is calculated as follows (Han et al., 2018; Heintzelman et al., 2002):

$$E_{trans}(k, d) = E_{elect}(k) + E_{amp}(k, d) = k(E_{Elect} + \epsilon_{uw}d^2) \quad (10)$$

Similar to terrestrial networks, each sensor node wastes power to transmit data (E_{trans}). E_{amp} denotes the power wasted by the transmitter amplifier, and ϵ_{uw} indicates the energy lost by the transmitter amplifier. The power lost by the receiver nodes for packet delivery is given by the following:

$$E_{res}(k) = E_{elect}(k) = k(E_{Elect}) \tag{11}$$

In each round, energy is consumed by non-CH sensors to transmit data to the CH and wasted by CHs to collect the data, and finally, the AUV collects the data. To address this issue, we considered the range of communication (r) and height angle (φ) to calculate $h = r \tan(\varphi)$. This equation is used to calculate the optimal depth at which the AUV is expected to maximize communication and data collection. The power of the AUV should not be an issue, as it can be recharged or replaced as needed. Therefore, we assumed that the AUV will perform all clustering and CH selection operations.

4 Numerical results and discussion

The efficacy of our suggested algorithm was assessed using a UWSN employing MathWorks MATLAB R2020b. Network parameters were established in accordance with the system model presented at the beginning of the paper. Table 2 shows the simulation parameters. Each sensor’s energy level was 5 Joules (Heinzelman et al., 2002; Omeke et al., 2021). The acoustic communication range was fixed at 2 m. To ensure optimal performance by the transmitter, we set the minimum power required by the receiver to -90 dBm (Omeke et al., 2021). To evaluate the network dimensions, we proposed to examine the configuration of 200 sensory nodes scattered across an approximate expanse of 0.001 cubic kilometers in the aquatic environment. While considering the attributes of the UWSN and the limitations of the resources, these factors must be considered.

Table 2 Simulation parameters

Parameters	Values
Simulator	MATLAB R2020b
Initial nodes energy (J)	5
Number of nodes	200
Simulation area (m ³)	10 ⁶
Transmission range (m)	2
Interference range (m)	2
Network topology	Random
minimum power required by the receiver (dBm)	-90

After the dynamic ascertaining of the most suitable number of clusters for every iteration through the utilization of the elbow method, the k-means algorithm clusters the UWSN. We additionally established a predetermined threshold for the network’s reclustering process, which was to be triggered when the power level of the CH sensor reached a point corresponding to 95% of the remaining energy average of the non-CH nodes within the current cluster.

Figure 5 depicts the convergence behavior of the DRL CH election technique in our proposed UWSN system. The aforementioned figure elucidates the performance of the proposed methodology in terms of the overall reward. It also demonstrates the notable success in the enhancement of the Q-learning algorithm through the integration of prior knowledge to initialize Q-values and utilization of specific techniques to update Q-values, ultimately reducing the convergence time and improving the learning speed of the agent. The primary observation entailed the expeditious convergence of the proposed DRL clustering algorithm, which led to favorable outcomes between the 145th and 600th episodes of the proposed approach. Moreover, our approach acquired the optimal policy for acquiring positive rewards following 1 200 episodes and culminated in complete convergence after 800 episodes.

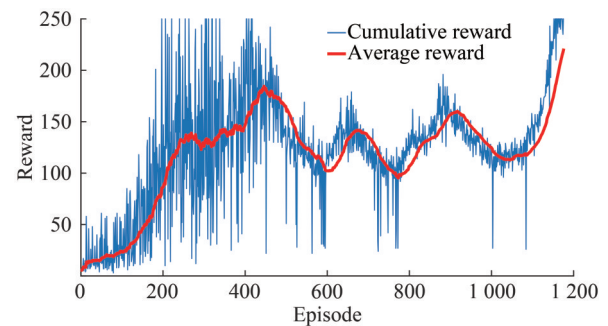


Figure 5 Convergence of the proposed DRL clustering algorithm by the proposed agent

Figure 6 illustrates a 3D network clustering and CH election using deep learning. The figure highlights the centroid of each cluster within the space by a “+” sign. Spatial analysis demonstrated the presence of clusters with dense deployment and others with sparse deployment.

Consequently, the DRL CH selection algorithm stated in

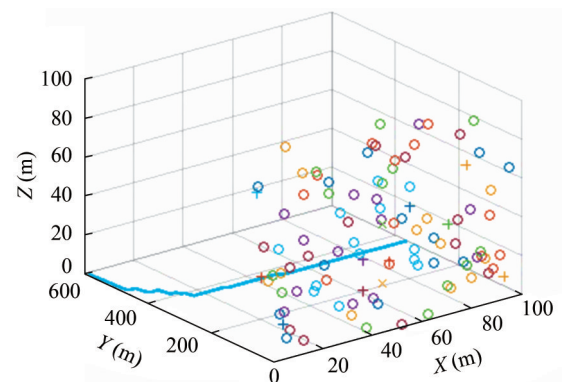


Figure 6 Distribution of network nodes in a 3D space. Nodes that coexist within a cluster and have been assigned a uniform color scheme. Moreover, the centroid of each cluster is demarcated with a (+) sign. In addition, the application of the DRL algorithm enabled the AUV to navigate toward the CHs along the blue curved path

Section 3 must consider two crucial factors. First, the protocol should exhibit a smaller Euclidean distance than the cluster centroid and other nodes within the cluster. Second, the protocol should ensure that the CHs' remaining energy is higher than that of other network nodes. At the end of every iteration (round), the AUV procures the data gathered by the CHs from non-CH nodes.

As expressed in the initial segment of this paper, the network's energy efficacy in terms of directly influenced the longevity of the UWSN, which can be assessed by considering the nodes with sustained functionality within the network. Figure 7 illustrates the number of functional nodes during each simulation round for four different algorithms in comparison with the proposed technique. The criterion for termination was achieved when a cluster devoid of nodes fulfilled the stated prerequisites to be appointed as a CH.

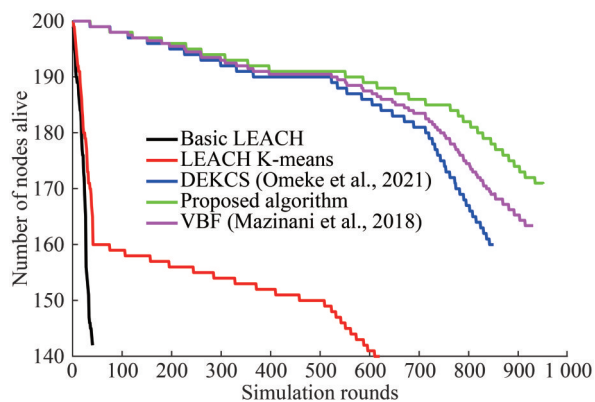


Figure 7 Plot showing the number of nodes that remain operational despite network disconnection

The aforementioned diagram illustrates the effect of four policies on network energy efficiency. Our proposed DRL algorithm successfully maintains a higher number of active nodes per round compared with other algorithms. Although the standard LEACH algorithm terminates in less than 50 simulation rounds, it is notably surpassed by LEACH k-Means and DEKCS, which operate for 640 and 840 rounds, respectively. The VBF protocol utilizes location data and VBF to determine the subsequent packet delivery hop. As a result, the VBF routing protocol faces a challenge as nodes constantly change positions, which leads to difficulties in maintaining stable routing paths throughout the network's lifespan. As a result, a minimal performance difference was observed between the VBF and our proposed method. The adaptive decision-based nature of our proposed method when considering real-time environmental data, node energy levels, and network conditions leads to a more efficient resource utilization per round, which results in underscoring its effectiveness in practical scenarios.

An examination of the area beneath the plotted line

depicted in Figure 8 proved that the recommended algorithm exceeds the LEACH algorithm by a substantial margin. Correspondingly, our suggested algorithm also surpasses the enhanced K-means LEACH, the DEKCS algorithm, and the VBF protocol. This difference in performance can be attributed to the arbitrary CH selection of the LEACH algorithm without considering their geographical position within the network or their remaining energy level. On the contrary, the utilization of the LEACH k-means algorithm necessitates the identification of nodes in close proximity to the cluster's centroid as CHs. However, the algorithm frequently neglects the remaining energy of nodes in most implementations. Our approach utilizes the elbow technique to identify the most suitable number of clusters and employs the DRL-based process for CH selection to include more nodes within the CHs' cluster zones, which reduces power consumption and conserves energy. Moreover, given the dynamic nature of the network and the need for periodic reclustering, the elbow technique guarantees optimal network fragmentation, which is a critical aspect in UWSN due to the harsh environmental conditions. The location of the CHs was determined well to guarantee a lower power consumption by the non-CH node amplifier. Furthermore, by training the DNN over time, the selection of more efficient CHs becomes highly probable.

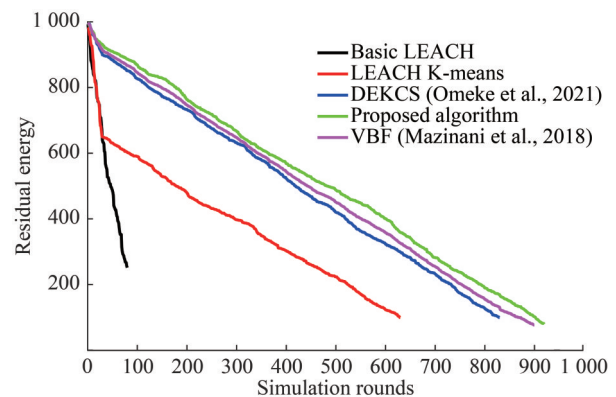


Figure 8 Plot presenting the remaining power of sensors in the network

As shown in Figure 8, our proposed DRL clustering method effectively protracted the lifespan of the UWSN by guaranteeing the complete utilization of residual energy in the network's nodes prior to network failure. Such an outcome was observed because the method considers the energy consumption weight of each CH node during each cycle of the network's lifetime, which resulted in a more balanced energy distribution among nodes during data transmission. DRL has the potential to offer more adaptive and energy-efficient solutions in dynamic underwater environments, where protocols such as VBF may face challenges due to periodic updates and adjustments to maintain efficient routing paths. It can also continuously learn and

adapt its CH election decisions based on changing conditions. This condition explains why DRL-based protocols represent a more advanced and potentially more efficient approach to energy consumption optimization in underwater sensor networks. CH nodes frequently transmit data directly to the AUV, which hovers at a certain depth in the water column. This condition results in a remarkable reduction in the energy consumption of the CH and non-CH nodes, which ultimately leads to a lesser depletion of energy.

Visual representations are frequently presented to demonstrate the death of sensors within a network. In Figure 9, the number of nonfunctioning nodes demonstrates that our suggested approach outperformed other algorithms under similar network circumstances, which enabled the network to operate for a considerably longer time with a reduced number of nodes ceasing to function. Basic LEACH and LEACH-K-means exhibited resemblances in their cluster-oriented approach toward energy efficiency but encountered difficulties in addressing inactive nodes due to their static or initial formation strategies. This inadequacy can disrupt cluster establishment and affect network performance through imbalanced energy usage and decreased efficiency. Although both protocols engage in data aggregation and processing within clusters to reduce communication overhead and conserve energy, inactive nodes still pose a threat to network functionality. Factoring in the AUV's depth hovering within the network, the DEKCS protocol prioritizes energy efficiency and cluster supervision to mitigate the presence of inactive nodes by electing optimal CHs and conserving energy during data transmission. Inactive nodes within the VBF can potentially interfere with routing paths and communication links, which results in challenges in the efficient management of inactive nodes as it depends on predetermined routing paths rooted in geographic data. This condition leads to suboptimal routing choices, increased retransmissions, and probable network segmentation. The proposed DRL technique exhibits improved efficacy in identifying node failures and triggering the timely initiation of new rounds and decluttering procedures. The utilization of AUVs for data collection optimizes data retrieval processes, diminishes energy usage, and increases network efficiency. This outcome empowers the network to sustain functionality for a markedly prolonged period while experiencing fewer node malfunctions.

The use of a trained DRL agent ensures stable floating operations per second (FLOPS) for scheduled loads, with the number of FLOPs being solely dependent on the network's architecture. DRL techniques were proposed based on the reliability of machine learning methods in uncertain UWSN environments. In the proposed algorithm, all neural networks comprised fully connected layer networks. The DQN model undergoes offline training, with the policy network being the only requirement for control actions post training. The number of FLOPS during inference is

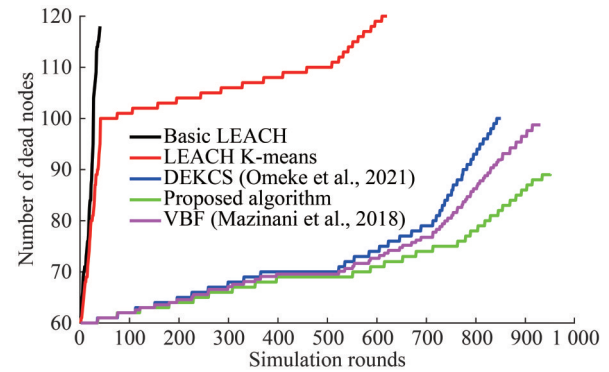


Figure 9 Plot showing the number of dead nodes despite network simulation rounds

predominantly determined by matrix multiplications of the activation policy network and the bios, which consists of three hidden layers with sizes of 64, 32 and 16, with N as the number of sensors serving as input to the neural network and C as the output and representing the number of CHs equivalent to that of clusters calculated via the elbow method. Thus, the FLOPS for the policy network during inference were computed as follows:

$$\text{FLOPS} = 64(N + 1) + 32(64 + 1) + 16(32 + 1) + C(16 + 1) = 64N + 17C + 2672 \quad (12)$$

The proposed approach's complexity is of linear polynomial order, with the total number of network iterations denoted as follows:

$$\mathcal{O}(I_{\text{iter}}(n)) \quad (13)$$

Furthermore, the suggested algorithm exhibits a linear time complexity, as stated in the aforementioned correlation. As the AUV manages all clustering and CH selection operations, its effect on network lifetime can be commented upon.

5 Conclusions

We proposed dynamic clustering in 3D-UWSNs through the utilization of PO-MD models and DRL algorithms to enhance energy efficiency and network lifetime. The adoption of the DRL protocol is essential in amplifying the scalability and efficiency of underwater systems over time due to its competency in learning and enhancement from experience. The precision and effectiveness of the network were enhanced through the joint implementation of the elbow method and k-means algorithm, which facilitated the identification of the optimal number of clusters and node locations within each cluster. This system showed promise in boosting the function of underwater systems. Our CH

election method considered residual energy, sensor coordinates, and Euclidean distance to designate the most appropriate CH in each network round, which ensured reliable transmission without additional communication overhead and employing a bias reward function to enhance routing efficiency, with data collection supported by an AUV. The selection criteria guaranteed that the nodes transmitted data with minimal power by locating the CH in close proximity, which resulted in enhanced network energy efficiency. Our performance evaluations manifested that the proposed protocol reduced the number of dead nodes per round and exploited more network energy than previous algorithms. Prospective research on DRL-based UWSN clustering protocols should focus on achieving equilibrium among coverage, delay, and reliability in relation to AUV-supported data collection while incorporating methods for faster convergence and increased learning speed of the agent through initialized and updated Q-values.

Competing interest The authors have no competing interests to declare that are relevant to the content of this article.

References

- Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: a survey. *Computer Networks* 38(4): 393-422. [https://doi.org/10.1016/S1389-1286\(01\)00302-4](https://doi.org/10.1016/S1389-1286(01)00302-4)
- Alsaman L, Alotaibi E (2021) A balanced routing protocol based on machine learning for underwater sensor networks. *IEEE Access* 9: 152082-152097. <https://doi.org/10.1109/ACCESS.2021.3126107>
- Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA (2017) Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine* 34(6): 26-38. <https://doi.org/10.1109/MSP.2017.2743240>
- Balhara S, Gupta N, Alkhatayy A, Bharti I, Malik RQ, Mahmood SN, Abedi F (2022) A survey on deep reinforcement learning architectures, applications and emerging trends. *IET Communications* 2022: 1-16. <https://doi.org/10.1049/cmu2.12447>
- Bhattacharjya K, Alam S, De D (2019) CUWSN: energy efficient routing protocol selection for cluster based underwater wireless sensor network. *Microsystem Technologies* 28(2): 1-17. <https://doi.org/10.1007/s00542-019-04583-0>
- Bholowalia P, Kumar A (2014) EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications* 105(9): 17-24
- Das D, Ameer PM (2017) Energy efficient geographic clustered multi-hop routing for underwater sensor networks. *TENCON 2017-2017 IEEE Region 10 Conference*, 409-414. <https://doi.org/10.1109/TENCON.2017.8227899>
- Fattah S, Gani A, Ahmedy I, Idris MYI, Targio Hashem IA (2020) A survey on underwater wireless sensor networks: requirements, taxonomy, recent advances, and open research challenges. *Sensors* 20(18): 5393. <https://doi.org/10.3390/s20185393>
- Gupta S, Singh NP (2024) Underwater wireless sensor networks: a review of routing protocols, taxonomy, and future directions. *The Journal of Supercomputing* 80(4): 5163-5196. <https://doi.org/10.1007/s11227-023-05646-w>
- Han G, Shen S, Song H, Yang T, Zhang W (2018) A stratification-based data collection scheme in underwater acoustic sensor networks. *IEEE Transactions on Vehicular Technology* 67(11): 10671-10682. <https://doi.org/10.1109/TVT.2018.2867021>
- Heinzelman WB, Chandrakasan AP, Balakrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications* 1(4): 660-670. <https://doi.org/10.1109/TWC.2002.804190>
- Hu T, Fei Y (2010) QELAR: A machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks. *IEEE Transactions on Mobile Computing* 9(6): 796-809. <https://doi.org/10.1109/TMC.2010.28>
- Javaid N, Hafeez T, Wadud Z, Alrajeh N, Alabed MS, Guizani N (2017) Establishing a cooperation-based and void node avoiding energy-efficient underwater WSN for a cloud. *IEEE Access* 5: 11582-11593. <https://doi.org/10.1109/ACCESS.2017.2707531>
- Jawhar I, Mohamed N, Al-Jaroodi J, Zhang S (2018) An architecture for using autonomous underwater vehicles in wireless sensor networks for underwater pipeline monitoring. *IEEE Transactions on Industrial Informatics* 15(3): 1329-1340. <https://doi.org/10.1109/TII.2018.2848290>
- Khalid M, Ullah Z, Ahmad N, Arshad M, Jan B, Cao Y, Adnan A (2017) A survey of routing issues and associated protocols in underwater wireless sensor networks. *Journal of Sensors* 7539751: 1-17. <https://doi.org/10.1155/2017/7539751>
- Khan A, Javaid N, Latif G, Jatta L, Fatima A, Khan W (2018) Cluster based and adaptive power controlled routing protocol for underwater wireless sensor networks. *21st Saudi Computer Society National Computer Conference (NCC)*, 1-6. <https://doi.org/10.1109/NCG.2018.8592955>
- Khan G, Gola KK, Ali W (2015) Energy efficient routing algorithm for UWSN-A clustering approach. *Second International Conference on Advances in Computing and Communication Engineering*, 150-155. <https://doi.org/10.1109/ICACCE.2015.42>
- Khan W, Wang H, Anwar MS, Ayaz M, Ahmad S, Ullah I (2019) A multi-layer cluster based energy efficient routing scheme for UWSNs. *IEEE Access* 7: 77398-77410. <https://doi.org/10.1109/ACCESS.2019.2922060>
- Khisa S, Moh S (2021) Survey on recent advancements in energy-efficient routing protocols for underwater wireless sensor networks. *IEEE Access* 9: 55045-55062. <https://doi.org/10.1109/ACCESS.2021.3071490>
- Kozłowski M, McConville R, Santos-Rodríguez R, Piechocki R (2018) Energy efficiency in reinforcement learning for wireless sensor networks. *ArXiv Preprint ArXiv: 1812.02538*. <https://doi.org/10.48550/arXiv.1812.02538>
- Li Y (2017) Deep reinforcement learning: An overview. *ArXiv Preprint ArXiv: 1701.07274*. <https://doi.org/10.48550/arXiv.1701.07274>
- Lu Y, He R, Chen X, Lin B, Yu C (2020) Energy-efficient depth-based opportunistic routing with q-learning for underwater wireless sensor networks. *Sensors* 20(4): 1025. <https://doi.org/10.3390/s20041025>
- Luo J, Chen Y, Wu M, Yang Y (2021) A survey of routing protocols for underwater wireless sensor networks. *IEEE Communications Surveys & Tutorials* 23(1): 137-160. <https://doi.org/10.1109/COMST.2020.3048190>
- Majid A, Azam I, Waheed A, Zain-ul-Abidin M, Hafeez T, Khan ZA, Qasim U, Javaid N (2016) An energy efficient and balanced energy consumption cluster based routing protocol for underwater wireless sensor networks. *IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, 324-333. <https://doi.org/10.1109/AINA.2016.165>

- Mazinani SM, Yousefi H, Mirzaie M (2018) A vector-based routing protocol in underwater wireless sensor networks. *Wireless Personal Communications* 100(4): 1569-1583. <https://doi.org/10.1007/s11277-018-5654-0>
- Nasir YS, Guo D (2019) Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks. *IEEE Journal on Selected Areas in Communications* 37(10): 2239-2250. <https://doi.org/10.1109/JSAC.2019.2933973>
- Nazareth P, Chandavarkar BR (2024) Cluster-based multi-attribute routing protocol for underwater acoustic sensor networks. *Wireless Personal Communications* 134(2): 781-808. <https://doi.org/10.1007/s11277-024-10926-6>
- Nguyen NT, Le TTT, Nguyen HH, Voznak M (2021) Energy-efficient clustering multi-hop routing protocol in a UWSN. *Sensors* 21(2): 627. <https://doi.org/10.3390/s21020627>
- Omeke KG, Mollel MS, Ozturk M, Ansari S, Zhang L, Abbasi QH, Imran MA (2021) DEKCS: A dynamic clustering protocol to prolong underwater sensor networks. *IEEE Sensors Journal* 21(7): 9457-9464. <https://doi.org/10.1109/JSEN.2021.3054943>
- Rahmani M, Dehghani MJ, Xiao P, Bashar M, Debbah M (2022) Multi-agent reinforcement learning-based pilot assignment for cell-free massive MIMO systems. *IEEE Access* 10: 120492-120502. <https://doi.org/10.1109/ACCESS.2022.3221935>
- Rowshanrad S, Keshtgary M, Javidan R (2014) MBC: A multi-hop balanced clustering routing protocol for wireless sensor networks. *International Journal of Artificial Intelligence and Mechatronics* 2(6): 164-170
- Shelar PA, Mahalle PN, Shinde GR, Bhapkar HR, Tefera MA (2022) Performance-aware green algorithm for clustering of underwater wireless sensor network based on optical signal-to-noise ratio. *Mathematical Problems in Engineering* 2022: 1-18. <https://doi.org/10.1155/2022/1647028>
- Souiki S, Hadjila M, Feham M (2015) Fuzzy based clustering and energy efficient routing for underwater wireless sensor networks. *Int. J. of Comp. Netw. & Commun. (IJCNC)* 7(2): 33-44. <https://doi.org/10.5121/ijcnc.2015.7203>
- Stojanovic M, Preisig J (2009) Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE Communications Magazine* 47(1): 84-89. <https://doi.org/10.1109/MCOM.2009.4752682>
- Sun Y, Zheng M, Han X, Li S, Yin J (2022) Adaptive clustering routing protocol for underwater sensor networks. *Ad Hoc Networks* 136: 102953. <https://doi.org/10.1016/j.adhoc.2022.102953>
- Sutton RS, Barto AG (2018) Reinforcement learning: An introduction. MIT Press
- Taheri R, Parsaei MR, Javidan R (2016) A new method for optimizing energy consumption in wireless sensor networks using enhanced LEACH protocol. *Journal of Engineering and Applied Sciences* 100(3): 576-581. <https://doi.org/10.3923/jeasci.2016.576.581>
- Wang C, Shen X, Wang H, Zhang H, Mei H (2023) Reinforcement learning-based opportunistic routing protocol using depth information for energy-efficient underwater wireless sensor networks. *IEEE Sensors Journal* 23(15): 17771-17783. <https://doi.org/10.1109/JSEN.2023.3285751>
- Wang K, Gao H, Xu X, Jiang J, Yue D (2015) An energy-efficient reliable data transmission scheme for complex environmental monitoring in underwater acoustic sensor networks. *IEEE Sensors Journal* 16(11): 4051-4062. DOI: 10.1109/JSEN.2015.2428712
- Wang M, Chen Y, Sun X, Xiao F, Xu X (2020) Node energy consumption balanced multi-hop transmission for underwater acoustic sensor networks based on clustering algorithm. *IEEE Access* 8: 191231-191241. <https://doi.org/10.1109/ACCESS.2020.3032019>
- Zhu F, Wei J (2018) An energy efficient routing protocol based on layers and unequal clusters in underwater wireless sensor networks. *Journal of Sensors* 2018: 1-10. <https://doi.org/10.1155/2018/5835730>