Iraqi Journal for Computers and Informatics
Available online: https://ijci.uoitc.edu.iq/index.php/ijci
Volume 52, issue 1. 2026

Print ISSN: 2313-190X, Online ISSN: 2520-4912

Hybrid Deep Reinforcement Learning and Bio-Inspired Optimization for Adaptive Routing and Clustering in Wireless Sensor Networks

Basim Jamil Ali [a]

[a] *Coputer Science Department, College of Science, Mustansiriyah University, Baghdad, Iraq.*

ARTICLEINFO

Keywords:

Deep Reinforcement Learning, Whale Optimization Algorithm, Clustering, Routing, Wireless

ABSTRACT

Wireless Sensor Networks WSNs are widely adopted and cost-effective way of implementing intelligent solutions in many low-resource settings. Both known Clustering/Routing Protocols such as Low-Energy Adaptive Clustering Hierarchy LEACH and Threshold-Sensitive Energy-Efficient Network Protocol TEEN as well as other related methods such as WOA-LEACH, are severely limited by their inability to support a long term network cycle for use in real world applications. This is because they all experience rapid energy exhaustion because of their inability to adapt to varying levels of energy and traffic demand. In addition, failure to distribute energy usage evenly throughout the network. In order to resolve these issues, we will develop a new Hybrid Approach that combines the Whale Optimization Algorithm WOA for determining energy aware cluster heads, and Deep Reinforcement Learning DRL for providing an adaptive multi-hop routing protocol. The proposed hybrid DRL-WOA solution will make joint optimizations of cluster heads and routing nodes to determine optimized routes to minimize energy use while maximizing energy efficiency through the optimization of hop distances, thereby creating longer lasting and more reliable communication processes. Results from simulations run on a 100 node WSN environment demonstrate the hybrid DRL-WOA solution achieves better performance than LEACH, TEEN, WOA-LEACH and a DQN-based only routing solution, including 22% less total energy consumption, 60% extended First Node Death FND, and PDR improvements of 5-15% in comparison to each of the above mentioned base line protocols. All in all, the experimental results clearly demonstrate that the proposed Hybrid DRL-WOA approach leads to a considerable improvement of the energy efficiency, network lifetime and the reliability of data delivery of static WSNs.

1. INTRODUCTION

Wireless Sensor Networks WSNs comprise many low power wireless sensor nodes that collectively collect, analyze and send data to a central or main station via multi-hop wireless communication. Because of their capabilities for providing continuous and real time monitoring, WSNs have gained wide acceptance in the monitoring of resource constrained environments including environmental monitoring, industrial automation, and smart infrastructure applications [1-3]. However, despite the widespread use of WSNs, they are severely limited by several factors related to their operation, including; low battery life, computing limitations, memory constraints, and distance limitations associated with radio frequency transmission that negatively affect network longevity and overall reliability [4-7]. During regular operational conditions of a WSN, WSN's experience with various problems such as, unevenly distributed power consumption on each node, excessive traffic congestion at heavily trafficked

E-mail address: basimja6nd@uomustansiriyah.edu.iq. [a]
Corresponding* : Basim Jamil Ali
Received 23 January 2026, Accepted 6th March 2026

DOI: 10.25195/ijci.v52i1.726.

locations, wireless interference, unstable links, and environmental disturbances greatly impact network routing efficiency [8–10]. The combination of these issues further impact network communication reliability and contribute to early failure of individual nodes particularly during conditions of unevenly distributed traffic loads and power consumption throughout the network. Therefore, inefficiencies in routing and selecting cluster heads may create localized traffic congestion, increase packet loss and decrease overall network throughput. Many different types of clustering and routing strategies that can help overcome these challenges exist in the literature, such as probabilistic clustering and routing with thresholds, bio-inspired optimizations, and learning based solutions [11–14]. While many of these solutions improve certain aspects of performance, many separate the process of clustering from the process of routing. Therefore, many times the two processes are made independently, and therefore do not take each other's needs into account when making decisions. For example, selecting an energy efficient cluster head does not always mean that the optimal path for multi-hop routing will be established. In addition, some of the learning based solutions that were developed introduce a significant amount of additional computation and/or require significant amounts of parameters to be optimized, which may not be feasible for resource constrained sensor nodes and large scale deployment [15–16]. The hybrid framework as proposed in this paper uses the Whale Optimization Algorithm WOA to select cluster heads on the basis of residual energy and the condition of the network; it also uses Deep Reinforcement Learning DRL to learn efficient routing paths that can be selected from multiple hops during operation by interacting with the network. This proposed hybrid framework has never been used before with the combination of bioinspired global optimization techniques (the WOA) combined with learning based adaptive decision making DRL techniques for simultaneous control over the clustering process and the routing process to obtain an equalization of energy use and improvement of the overall performance of the network over time. Although the network architecture is assumed to be stable after deployment, routing behavior remains dynamic due to variations in data types transmitted across the network, remaining node power, and policy adjustments based on machine learning. Therefore, a hybrid framework based on Deep Reinforced Learning DRL and Wolf Optimization Algorithm WOA is well-suited to operate efficiently under these dynamic conditions, making it applicable to large-scale, resource-constrained wireless sensor networks. Beginning with section 3, the rest of the paper will be structured as follows. Section 2 provides a review of previous work regarding both clustering and routing within Wireless Sensor Networks WSNs. In section 3 we describe our system model and how we formulate the problem to which we are trying to find an answer. We present the details of the proposed DRL-WOA framework in section 4. In section 5, we provide information about the structure of our simulations and the metrics that we use to measure their quality; then we report on the results of those simulations and analyze them in section 6. In the final section 7, we summarize the findings of this paper and discuss possible ways in which the research could continue.

2. RELATED WORK

The majority of the recent WSN routing-related research have been focused on increasing routing efficiency through better cluster-head selection methods and longer network lifetime. There are many differences among these studies with respect to: the specific optimization strategy; the level of adaptability in each study; the efficiency of each algorithm; and the computational cost of each method. The following is a summary of the most applicable studies with respect to their fundamental concepts, benefits, and limitations as well as references to other research methods described.

1. E-PEGASIS 2021

E-PEGASIS was developed by Sadhana et al. to improve upon the data transmission efficiency of PEGASIS. This was accomplished by implementing an average inter-node distance as the sequencing criteria for their method and employing a constant radio range for the base station communications with the rest of the sensor nodes, initiating serialization based on nearest edge calculations. Although this method does offer better transmission overhead and improved performance when compared to the original PEGASIS, it still employs constant radio ranges and static assumptions; therefore, it lacks flexibility to adapt to fluctuations in network density and varying levels of energy per node [17].

2. Aggregation Routing 2021

The authors Daanoun et al., studied and compared a Distributed Aggregation Routing Protocol DARP to Low-Energy Adaptive Clustering Hierarchy LEACH by using probabilistic cluster head rotation and one hop communication to the base station, not only in terms of how much control overhead is decreased, but also with regards to scalability issues and unevenly distributed energy usage. The authors demonstrated that the probabilistic selection of cluster heads based upon residual energy contributes to rapid energy consumption near the base station and therefore indicates that the LEACH family of protocols will not be suitable for large-scale, multi-hop deployment [18].

3. Fuzzy Clustering + PSO / AWOA 2022

The following is an example of a system (Ramya et al.) utilizing fuzzy clustering as part of the methodology, with optimization methods PSO / AWOA to optimize several variables such as residual energy, node centrality, node degree and BS

distance. The data collected by Ramya's experiments showed that fuzzy clustering with multi-parameter optimization improved routing performance, network delay and energy usage. Although fuzzy clustering is beneficial in terms of routing, it can be computationally expensive and therefore less likely to be implemented on low resource sensor nodes due to the multi-optimization stage [19].

4. K-IABC + CL-HHO 2023

Xue et al. developed an energy efficient clustering method based on an artificial bee colony-based k-medoids algorithm K-IABC, and combined this with a CL-HHO cross-layer routing protocol. Their results show significant gains in terms of Packet Loss Ratio PLR, throughput, delay reduction and network longevity. While the dual-optimization framework provides good results, it also increases the complexity of the algorithm and can potentially increase response times when dealing with rapidly changing network dynamic [20].

5. MOD-LEACH

MOD-LEACH was first proposed by Devika et al. The authors incorporated an existing Butterfly Optimization Algorithm BOA with a deep reinforcement Learning DRL. They used this hybrid approach to optimize the tradeoff for BOA between exploration and exploitation while enabling DRL to dynamically decide how routes are established for data transmission in WSNs. Using MOD-LEACH instead of traditional MOD-LEACH resulted in approximately 22% less energy consumption than its predecessor. Although the method proposed by Devika et al. demonstrates improvements in performance, the use of neural network-based inference may degrade runtime efficiency due to the high power requirements of most WSN devices [21].

6. Swarm-Intelligence Clustering (ACO, PSO, BA, FA) 2024

Nainwal et al., compared swarming algorithms for their efficiency in clustering when there are limited amounts of energy available. They demonstrated that the ACO and PSO algorithms were relatively good at clustering under these conditions; however, as the size of the swarm increases so does the computational load and potentially the sensitivity of parameters used to describe the swarm's behavior. As stated above, none of the swarming algorithms described here, include an inherent mechanism for routing adaptability beyond selecting a CH [22].

7. ARL-DARO 2024

Shobana et al., developed a Q-learning based routing using an algorithm called ARL-DARO which dynamically aggregates data and refines it at runtime to eliminate redundancy and provide improved routing security and stability. Results from Shobana et al., indicate their use of the ARL-DARO algorithm resulted in a range of 20% to 45% reduction in overall energy usage and a range of 10% to 30% increase in data throughput. The ARL-DARO algorithms performance is very similar to all other applications of Q-learning and its performance will directly depend on the accuracy of the reward definition and design into the state space, and thus, will not be able to perform well when subjected to varying traffic conditions [23].

8. SHO-CH 2025

Prakash et al. presented SHO-CH using Spotted Hyena Optimization for cluster-head selection in heterogeneous WSNs. The approach showed substantial improvements over PSOBS, GAOC, PSO-ECSM, and NEHCP, achieving increases of 46.21% in settling time, 43.34% in network lifetime, and 49.46% in throughput. Yet, the heavy reliance on meta-heuristic computation requires tuning and may be slow at scale when node count increases significantly [24].

Table 1 indicates the comparison of almost all prior approaches and clearly demonstrates that nearly all prior approaches have individually optimized clustering and routing and thus, there was limited cooperation between optimizing energy aware cluster heads and optimizing adaptive multi-hop routes. Unlike previous works that treat clustering and routing sequentially/independently, this paper proposes a jointly optimized framework DRL–WOA for both the clustering process and routing process through a single learning-based framework and therefore achieves coordinated network wide performance improvements at low computational cost.

Table 1: *Comparative analysis related works with the proposed DRL–WOA framework*

Ref.	Approach	Clustering Method	Routing Method	Joint Optimization	Adaptability	Energy Balancing	Computational Cost
[17]	ARL-DARO	Heuristic / static clustering	Q-learning–based routing	✗	High	Limited	Medium
[18]	LEACH-based survey / enhancement	Probabilistic CH rotation	Single-hop / aggregation	✗	Low	Poor	Low
[19]	Fuzzy + PSO/WOA	Fuzzy multi-parameter clustering	Heuristic routing	✗	Medium	Moderate	High
[20]	K-IABC + CL-HHO	Bee colony–based k-medoids	Cross-layer heuristic routing	Partial	Medium	High	High
[21]	MOD-LEACH (DRL + BOA)	BOA-assisted CH selection	DRL-based routing	Partial	High	Moderate	High
[22]	Swarm-based (ACO/PSO/BA /FA)	Swarm intelligence clustering	Static routing	✗	Low	Moderate	Medium–High
[23]	ARL-based aggregation routing	Static / heuristic clustering	RL-based adaptive routing	✗	High	Limited	Medium
[24]	SHO-CH	Spotted Hyena Optimization	Static routing	✗	Low	High	High
—	**Proposed DRL–WOA**	**WOA-based energy-aware clustering**	**DRL-based adaptive multi-hop routing**	✓	**High**	**High**	**Moderate**

3. RESEARCH MOTIVATION AND SYSTEM MODEL

Most current approaches to the efficient clustering and routing of energy in Wireless Sensor Networks WSN consider the selection of the Cluster-Head and the Routing Process as separate Optimization Problems; therefore, they can lead to an unbalanced energy consumption and consequently to a Reduced Network Lifetime [19-22]. In addition, the majority of bio-inspiration based approaches are focused on the Clustering Efficiency, while most Reinforcement Learning (RL)-based approaches have been focused on the Adaptability of the Routing Process without sufficient Coordination with the Cluster Formation Process [23, 24].

Therefore, the main motivation of this study is to propose a Unified Hybrid Framework for Deep Reinforcement Learning DRL and Whale Optimization Algorithm WOA, which Jointly Optimizes Energy Aware Cluster Head Selection and Adaptive Multi-Hop Routing to Increase the Network Lifetime and Data Delivery Reliability in Static Deployments of Wireless Sensor Networks WSNs.

The following presents the complete system model used in the design and evaluation of the proposed framework. We consider a wireless sensor network deployed over a two-dimensional sensing field, consisting of a set of static nodes with limited battery energy and short-range wireless communication capabilities. The nodes periodically generate sensing data and forward it through multi-hop communication to a stationary base station. The model assumes heterogeneous energy conditions, where energy consumption depends on both transmission distance and packet forwarding activity.

We adopt the widely used first-order radio energy model to characterize communication cost, in which energy dissipation for packet transmission and reception is determined by amplifier and electronic circuitry parameters.

Residual energy is continuously tracked to capture battery depletion over successive communication rounds. In addition, node density, spatial distribution, and proximity to the base station are considered critical factors influencing cluster-head selection and routing reliability. The core optimization problem is to maximize network lifetime and packet delivery performance while minimizing overall energy expenditure and local congestion around heavily used relays. To this end, clustering and routing are treated as coupled optimization processes: WOA is responsible for selecting energy-balanced cluster heads, while the DRL agent learns adaptive multi-hop forwarding strategies based on real-time network conditions, including neighbor availability, link quality, and remaining energy.

3.1. Network Assumptions

The wireless sensor network is modeled as a graph G = (V, E), where:

- V = {v1, v2, ..., vN} represents the set of N sensor nodes.
- E denotes the set of wireless communication links based on node proximity and communication range.
- The key assumptions of the network are as follows:
  - Nodes are randomly and uniformly deployed in a 2D field of size L X L meters.
  - Each sensor node is equipped with a non-rechargeable battery and possesses limited processing and transmission capabilities.
  - A static Base Station BS is located either within or outside the sensing field.
  - Nodes periodically sense environmental data and transmit it to the BS via single-hop or multi-hop communication.
  - Nodes are stationary after deployment.
  - All nodes are aware of their own residual energy and position (either through localization or GPS).

3.2. Energy Consumption Model

Equations 1 and 2 explain the First-Order Radio Model for the energy consumption model for a k-bit message over a distance d. Transmission energy equation:

$$E_{TX}(k,d) = \{k.E_{elect} + k.\varepsilon_{fs}.d^2 \quad if\ d < d_0\ k.E_{elect} + k.\varepsilon_{mp}.d^4 \quad if\ d \geq d_0 \qquad (1)$$

where:

- k: size of the transmitted packet in bits.
- d: distance between sender and receiver.
- E_{elect}: energy consumed by electronics to transmit or receive 1 bit.
- ε_{fs}: amplifier energy for the free-space model with path-loss exponent n = 2.
- ε_{mp}: amplifier energy for multipath model with path-loss exponent n = 4.
- d_0: adopted threshold distance ($d_0 = sqrt(\frac{\varepsilon_{fs}}{\varepsilon_{mp}})$).

Reception energy equation:

$$E_{RX}(k) = k.E_{elect} \qquad (2)$$

The total energy consumed by a node includes sensing, processing, transmission, and reception, but the communication cost dominates and is the primary focus for optimization.

The hybrid system operates in two main stages per communication round:

A. Clustering Stage: WOA algorithm selects an optimal set of cluster heads based on a multi-objective fitness function considering: Residual energy of nodes, average intra-cluster distance, node centrality and density, and balancing CHs across the field.

B. Routing Stage: A Deep Reinforcement Learning DRL agent determines the most energy-efficient and reliable paths for: Intra-cluster communication (from member nodes to CH), and Inter-cluster (CH-to-BS) routing, possibly via multi-hop CH forwarding. The DRL agent continuously updates its policy by using network states such as energy levels, queue size, and hop count to BS, thereby learning optimal routing decisions over time.

3.3. Problem Formulation

Multi-objective functions are adopted in this model:

Objective 1: Prolong Network Lifetime

$$Maximize \quad R_{max} = total\ number\ of\ rounds\ before\ all\ nodes\ die$$

Objective 2: Minimize Energy Consumption per Round

$$Minimize\ E_{round} = \sum_{i=1}^{N} E_i^{used}(r) \qquad (3)$$

Where $E_i^{used}(r)$ is the energy consumed by node i during round r.

Objective 3: Maximize Packet Delivery Ratio PDR

$$Maximize\ \ PDR = \frac{Total\ packets\ received\ at\ BS}{Total\ packets\ generated} \qquad (4)$$

Objective 4: Ensure Load Balancing

Prevent premature death of any particular node by equalizing energy consumption among nodes through:

- Dynamic CH rotation and routing path diversity via learned DRL policy.
  - ✓ The binary decision variable aij is an indicator ($a_{ij} \in \{0, 1\}$) of the DRL agent's action choice process, where $a_{ij} = 1$ means that node i will choose node j as its next hop to forward packets, and $a_{ij} = 0$ otherwise. The DRL agent does not directly optimize a_{ij}. Instead, a_{ij} is generated through the routing policy learned by the DRL agent based on the current network state.
  - ✓ DRL action vector A_t: next-hop selection based on observed state S_t.
- Constraints:
  - ✓ $\sum CH_i \leq Max\ CH_s$
  - ✓ $E_i(t) > 0$: node must have energy to participate.
  - ✓ Nodes must be within communication range.

***Table 2**: Symbol Definitions*

Symbol	Description
(E_i)	Residual energy of node *i*
$\boldsymbol{E_{round}}$	Energy Consumption per Round
(N_i)	Neighbor list of node *i*
(a_t)	Selected next-hop action at time *t*
(s_t)	State vector observed at time *t*
$\boldsymbol{\alpha, \beta, \gamma}$	variable weights of rewards

The neighbor list is represented as an index-based discrete action space. Each possible action represents the choice of an available neighbor as the next hop.

- Clustering: Solved using bio-inspired metaheuristic with a multi-objective fitness function.
- Routing: Learned via DRL agent trained to optimize reward function defined as:

$$R_t = \alpha.\frac{1}{E_{tx}} + \beta.PDR - \gamma.Latency \qquad (5)$$

where:

α, β, γ are variable weights of rewards.

The same reward function is applied to all routing decision optimizations during the entire training process of the DRL to guarantee learning stability and the convergence of the training process. The optimization of all routing decisions is based on one common reward function definition .

The above-mentioned integrated approach will enable the system to be flexible in its response to dynamic network conditions; to manage energy consumption; to provide an intelligent routing decision process that will increase the life span of the network and the reliability of communications.

4. THE PROPOSED HYBRID APPROACH

The use of the Whale Optimization Algorithm WOA, based on biological inspiration for a global search mechanism, has been explored in order to find the optimum Cluster Head CH assignment; WOA uses a randomly generated set of initial candidate locations for the whales in the solution space. During the exploration stage, the whales move around the solution space evaluating different CH configurations. The evaluation results in the selection of the most efficient energy configuration as the new "best" solution. Subsequently. The encircling and spiral update phases are implemented to refine the CH locations within the region identified from the previous phase by balancing between contraction behavior and logarithmic spiral movement to ensure that both convergence and diversity exist. The process continues until the solution converges or the iteration limit has been reached, at which time the best CH configuration will be selected for the next operational cycle.

4.1. WOA Workflow

The cluster head within the Whale Optimization Algorithm WOA is the optimal distribution among all possible distributions of cluster heads in the search space, by using the cooperative foraging strategy used by humpback whales to attain the best possible distribution. Each whale configuration is iteratively improved in order to find either a stable distribution that allows CH to remain stable, or else when the number of iterations specified as an iteration limit are reached. WOA uses an iterative process that includes both global exploration and local exploitation phases to refine the configuration of the whales from the initial random population of whales. When the whales perform global exploration, they can be randomly moved around to explore different distributions of CHs and to avoid premature convergence of solutions. As soon as a promising configuration is identified, the algorithm switches to local exploitation where it uses two different strategies to refine its search; a linear contraction strategy which contracts the position of the whales towards the CH distribution that has been found to be the most promising so far and a spiral updating strategy which updates the whales' positions along a logarithmic curve. It achieves the goal of maximizing both the diversity of the searches and the quality of the results from the search process. In that case the optimized placement of CH is the output and will provide the best distribution of energy loads; the shortest communication distances; and most importantly the longest-term routing stability for the network. A representation of the above method is shown below as pseudocode of the integrated WOA-DRL methodology used within this study.

```
Algorithm Hybrid WOA–DRL
1. Initialize all sensor nodes and assign initial energy
2. Generate whale population (CH candidates)
3. For i = 1 to max_Iterations do
     Evaluate fitness for each candidate using Eq. 6.
     Update best CH solution
     If |a| < 1 then
        Exploitation using encircling or spiral movement
              Else
        Exploration using random repositioning
   End For
4. Assign final CH configuration
5. For each transmission round do
    For each node do
        Observe state s (Eq. 7)
        Select action a (next hop)
        Send packet
        Receive reward r (Eq. 8)
        Update policy (Q or gradient) (Eq. 9)
     End For
  End For
6.Terminate when all nodes are depleted or transmission complete
```

4.2. Computational and Memory Complexity

The computational cost of the proposed hybrid DRL–WOA protocol is governed by the clustering and routing phases. During clustering, the Whale Optimization Algorithm performs iterative position updates of candidate solutions to determine the optimal cluster heads. Given a network with n nodes and i optimization iterations, the time complexity of this stage is approximately O(n · i), as each iteration requires evaluating the fitness of all candidate CH sets. In the routing phase, the DRL agent updates its policy based on feedback from the network environment. The computational complexity of DRL component can be expressed as O ($|S|\cdot|A|$), where the state space $|S|$ is proportional with the number of features per sensor node (e.g., residual energy, distance to sink or cluster head, buffer occupancy, and link quality) and the action space $|A|$ is limited by the maximum number

of neighboring nodes available for routing decisions. Consequently, a network with N nodes, each qualified by f features, and with a maximum neighbor degree d, the overall complexity can be recalled as $O((N \cdot f) \cdot d)$. This calculation effort grows in partnership with network size, feature wealth and connectivity, ensuring of the needing for hybrid models that decrease influence of search space while taking part adaptability. During the Clustering phase the memory usage is linear O(n), due to storage of node properties, fitness values etc. In addition to the memory requirements mentioned above, the DRL portion of the algorithm also requires additional memory for the replay buffer, temporary states from intermediate networks and parameters of the neural models; although this increased memory overhead should still be feasible for relatively small to medium sized networks. For larger deployments, the DRL learning may be performed off-line or at a central location so that they do not burden the sensor devices, and thus scalable.

4.3. Framework of the Proposed Hybrid DRL–WOA Protocol

The flow of the proposed Hybrid DRL-WOA Framework is shown in figure 1. The WOA is used to choose globally optimal energy-balanced cluster heads after the sensors have been deployed and their initial energies have been set. Cluster head selections are followed by the clustering of sensor nodes into their respective cluster heads for the collection of data . In every communication round, a DRL-based routing process is run. Each node views its current environment which includes its remaining energy level, the current queue status, what neighbors are available, and how far it is from the base station. With all these in view, using an epsilon-greedy policy, the next routing action (which will be the next-hop selection) is determined. Upon sending the packets, the rewards are then calculated that will take into account the amount of energy consumed while transmitting the packets successfully as well as the time delay, and the DRL policy will be updated as needed. The learning process of the framework iteratively continues over multiple rounds until there is no more energy left in the network. This allows the proposed framework to use the WOA algorithm to determine the cluster head of the sensor networks globally and optimally; and the DRL algorithm to route the data through the network in an adaptive manner resulting in the most efficient usage of energy and the longest possible lifetime of the network.

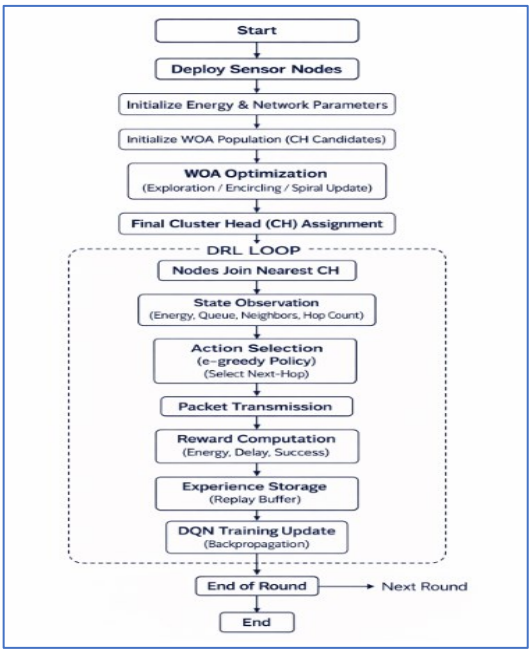

Fig. 1: *Framework of the proposed hybrid DRL–WOA protocol*

4.4. Overview of the Hybrid Framework

Each communication round in the network consists of the following sequence:

1. **Clustering Phase**: In this study, we apply the WOA to identify an optimal set of cluster heads that minimize intra-cluster communication cost and balance energy consumption across the network.

Objective Function:

The fitness function used in the optimization process is defined as a weighted combination of key metrics:

$$Fitness(i) = \omega_1 . \frac{E_i^{residul}}{E_{max}} - \omega_2 . \frac{D_i^{intra}}{D_{max}} - \omega_3 . \frac{N_i^{neighbors}}{N} \qquad (6)$$

Where:
$E_i^{residul}$: residual energy of node i,
D_i^{intra}: average distance to other nodes in its cluster,
$N_i^{neighbors}$: number of neighbors (node density),
$\omega_1, \omega_2, \omega_3$: weighting factors (tuned empirically)

In setting the coefficients for the clustering fitness function during grid search experiments, the weights ω_1, ω_2, and ω_3 which represent, respectively, the contributions from the residual energy, average distance, and node density were tuned empirically to 0.4, 0.35, and 0.25. These were derived to strike the best equilibrium among energy consumption uniformity and the stability of clustering.
The optimization algorithm searches the solution space to maximize this fitness function and select CHs accordingly.
The clustering phase outputs a binary vector indicating the CH nodes for the current round, which is passed to the DRL-based routing module.

2. **Routing Phase**: Deep Q-Network used as DRL agent learns to select the most energy-efficient and reliable routes for data transmission from sensor nodes to the Base Station BS, either directly or through intermediate CHs. Once clusters are formed, a DRL agent determines the best forwarding path from sensor nodes (or CHs) to the base station, adapting to real-time network conditions.

Each node's state at time t is defined by:
- Residual energy of the node.
- Queue size (buffer load).
- Distance to the base station.
- Neighboring node list and their energy levels.
- Number of hops to BS.

$$S_t = \{E_i^{res}, Q_i, D_i^{BS}, Neighbors_i, H_i\} \quad (7)$$

Action Space (A_t)
Possible actions correspond to the selection of the next-hop node from the current node's neighbor list.
$A_t = \{Select\ neighbor\ j \in Neighbors_i\}$

Reward Function (R_t)
The DRL agent is trained by using a reward function that balances energy consumption, reliability, and latency:

$$R_t = \alpha.\frac{1}{E_{TX}} + \beta.SuccessRate - \gamma.Delay \quad (8)$$

Where:
E_{TX}: energy used in transmission.
Success Rate: initially binary indicator of packet delivery, then in the reward formulation it is handled as a continuous variable, which coincides with a smoothed / averaged success rate over a sliding time window / across multiple transmission attempts.
Delay: estimated end-to-end delay
α, β, γ: reward weights.
The reward encourages selection of energy-efficient, high-success, and low-latency paths.

Learning Algorithm
By utilizing a Deep Q-Network DQN that approximates the Q-value function Q(St, At) by using a deep neural network.
The network is trained by using experience replay and a target network to improve stability.
Training follows the Bellman update:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \eta[R_t + \delta.Q(S_{t+1}, \grave{a}) - Q(S_t, A_t)\ ] \quad (9)$$

Where:
η: learning rate,
δ: discount factor for future rewards.

3. **Transmission Stage**: The data will then be sent through the chosen cluster heads and along the predetermined routes that were established by the DRL Agents' selections in this transmission stage. Simultaneously, the DRL agent will update the energy consumed by each node in the process of sending the data.

4. Combining Clustering and Routing

i. Beginning of a New Cycle:
ii. Cluster Development: Based upon current conditions of the network, an optimization method selects cluster heads.
iii. Policy Revision: The DRL Agent reviews the revised current condition of the network and modifies the routing choice about how to transmit information through the network.
iv. Data Exchange: Each node will send its messages to the pre-designated cluster-heads and follow the route that has been determined by the DRL Agent. The actions taken by the nodes will be logged by the DRL Agent in order to train the agent. This process repeats itself in each iteration, with the possibility of the network learning and adjusting continuously in order to find an optimal configuration for the network's topology (optimal cluster-head selection) and optimal routes (optimal routing path).

5. RESULTS FROM EXPERIMENTS AND PERFORMANCE ASSESSMENT

This area assesses the hybrid DRL-WOA methodology's performance through simulations on identical network environments in comparison to LEACH and TEEN, WOA-LEACH, and a DQN-only routing method for purposes of establishing a baseline of learning-based routing. The performance is measured based upon a variety of criteria such as network lifetime, total energy consumed by the network, Packet Delivery Ratio PDR and routing overhead.

5.1. Simulation Testbed

The simulation experiments were conducted in a Python testbed by using the TensorFlow + TF-Agents for handling the DRL model and NetworkX libraries. There were 100 sensor nodes placed randomly throughout an area of 100m x 100m. The location of the base station was (100, 50). The radio energy model was a simple first order model with Eelec = 50nJ / bit and a threshold distance d0 = approximately 87 meters. Table 3 contains a full list of the simulation parameters.

Table 3: *Simulation Parameters*

Parameter	Value
Platform	**Python 3.10 with TensorFlow and NetworkX.**
Simulation Area	**100 m × 100 m square field.**
Number of Sensor Nodes	**100 nodes (randomly deployed).**
Base Station Location	**Fixed at coordinates (100, 50), outside the sensor field**
Communication Model	**Single-hop intra-cluster, multi-hop inter-cluster.**
Initial Energy per Node	**2 Joules.**
Packet Size	**4000 bits.**
Data Aggregation Energy	**5 nJ/bit/signal.**
Radio Parameters	**Based on the first-order energy model:**
E_{elec}	**50 nJ/bit**
ε_{fs}	**10 pJ/bit/m^2**
ε_{mp}	**0.0013 pJ/bit/m^4**
d_0	**sqrt** $\left(\frac{\varepsilon_{fs}}{\varepsilon_{mp}}\right) \approx 87\ m$

5.2. Performance Metrics

Performance metrics were measuring how well the protocol performed. The following is a list of these metrics:

1. Network Lifetime:
   - First Node Dies FND: Round in which the first sensor node runs out of energy.
   - Half Nodes Die HND : Round at which 50% of the nodes are dead.
   - Last Node Dies LND: Round when the last node runs out of energy.
2. Total Energy Consumption: Energy amount of total energy that was used by each round, or amount of total energy that was used at any given point of time.
3. Packet Delivery Ratio PDR:

$$PDR = \frac{Total\ packets\ received\ at\ BS}{Total\ packets\ generated\ by\ all\ nodes\}} \qquad (10)$$

4. The overall, or average: The time it takes for all packets to go through the network, beginning at the source node, ending at the base station.
5. The ratio of overhead to data packets: The ratio of the total number of packets that have been used to send routing information (overhead) compared to the number of packets of data that were transmitted within the network.
6. Degree of CH Load Balancing: Standard deviation of energy consumption among all cluster heads.

5.3. Simulation

Each simulation will be run for 2000 rounds or until all nodes are dead. Each of the protocols will be running thirty independent runs with random deployments and initial selections of Cluster Head CH and the results will be averaged to provide a reliable estimate based on statistics.

5.4. DRL Training Paradigms

DRL Settings for the Deep Q-Network Component:

- State Features: Residual energy, neighbor energy, hop count, queue size.
- Action Space: Neighbor selection (next-hop).
- Replay Buffer Size: 10,000 experiences.
- Mini-Batch Size: 64.
- Learning Rate: 0.001.
- Discount Factor δ: 0.9.
- Exploration Strategy: Epsilon-greedy (ε starts at 1 and decays to 0.01).
- Neural Network: 3 layers with 64–128–64 neurons (ReLU activations).
- Training Interval: DRL agent is trained at the end of each round.

6. RESULTS AND DISCUSSION

To evaluate our proposed hybrid algorithm, we compare the simulation results from our proposed hybrid DRL-WOA algorithm with other baselines including: LEACH; TEEN; WOA-LEACH; and DQN-only. We Use four different metrics to evaluate the algorithms' performance: network lifetime; energy consumed; packet delivery ratio; and routing overhead. Table 4 provides a comprehensive overview for comparing the qualitative results of the proposed DRL-WOA to all of the baseline methods presented.

Table 4: *Comparative performance analysis of routing protocols*

Protocol	Energy Efficiency	Network Lifetime	PDR	Routing Adaptability	Optimization Strategy
LEACH	Low	Short	Low	Static	Heuristic clustering
TEEN	Medium	Medium	Medium	Event-driven	Threshold-based
WOA-LEACH	High	Medium	Medium	Static	Bio-inspired clustering
DQN-only	Medium	Medium	High	Adaptive	DRL routing only
Proposed DRL–WOA	High	Long	High	Adaptive	WOA + DRL (Unified)
LEACH	Low	Short	Low	Static	Heuristic clustering

The proposed DRL–WOA framework achieves balanced improvements across all key performance metrics by jointly optimizing cluster-head selection and adaptive multi-hop routing.

- **Network Lifetime**

The network lifetime is an important measure of the performance of a Wireless Sensor Network WSN and is generally assessed with 3 metrics; First Node Dead FND, Half Node Dead HND and Last Node Dead LND. FND represents the round number when the first sensor node is completely exhausted, HND represents the round number when 50% of the sensor nodes are no longer active and LND represents the round number when all the nodes have been exhausted and there is no longer any activity within the network. Together these 3 metrics represent how efficiently and effectively the network is utilizing its available energy resources and how well it is distributing its energy usage among its various nodes.

As illustrated in figure 2, the number of live sensor nodes over the rounds of simulation is depicted for each of the routing protocols that were tested. As indicated in this figure, the DRL-WOA framework that was developed in this research has maintained a significantly larger number of active nodes than LEACH, TEEN, WOA-LEACH and DQN-only protocols over the rounds of simulation. This pattern of behavior indicates that the proposed framework is delaying the failure of individual sensor nodes and improving the overall balance of their energy usage, both of which can be directly attributed to the WOA-based selection of the energy aware cluster-heads and the DRL-based learning of the multi-hop routing paths that will minimize the amount of energy consumed during communication.

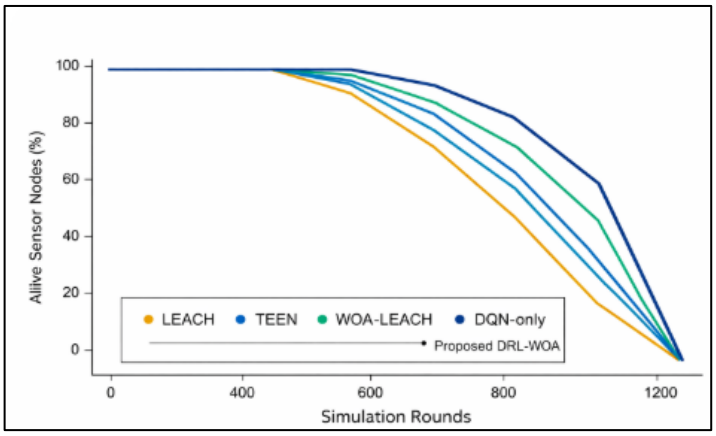

Fig. 2 : *Number of alive nodes versus simulation rounds for different routing protocols*

Table 5 shows the network lifetime between the different routing protocols. The table presents the FND, HND and LND values that were obtained through averaging the results of the multiple simulations. The DRL-WOA protocol that was developed in this research has achieved the longest network lifetime of all 3 metrics. To further quantify the gain, we see that performance of the proposed method in terms of percentage improvement (FND≅19.6%, HND≅15.4% and LND≅14.87%) over baseline (WOA-LEAH). Specifically, the large delays between the FND and HND rounds indicate that the proposed framework is preventing the premature exhaustion of the energy reserves of the individual nodes and that it is extending the LND round indicates that the proposed framework is also enhancing the overall sustainability of the network. Both of these benefits can be directly attributed to the joint optimization of the cluster-head rotation and the selection of the routing path that will distribute the communication load most evenly throughout the network.

Table 5: *Network lifetime comparison*

Protocol	FND (Rounds)	HND (Rounds)	LND (Rounds)
LEACH	**380**	**800**	**1040**
TEEN	**410**	**860**	**1100**
WOA-LEACH	**510**	**970**	**1210**
DQN-only	**470**	**910**	**1165**
Proposed Hybrid	**610**	**1120**	**1390**

Further validation is shown in figure 3 through the residual energy average for each round of simulation. Compared to all other methods examined as a base line, the DRL-WOA method always has greater residual energy. Furthermore, the difference between the two is increasing over time which validates that by jointly optimizing the cluster head selections based on their energy consumption and dynamically adjusting the number of hops used in multi-hop routing will lead to less energy being consumed in the network overall and therefore more stable networks over longer periods.

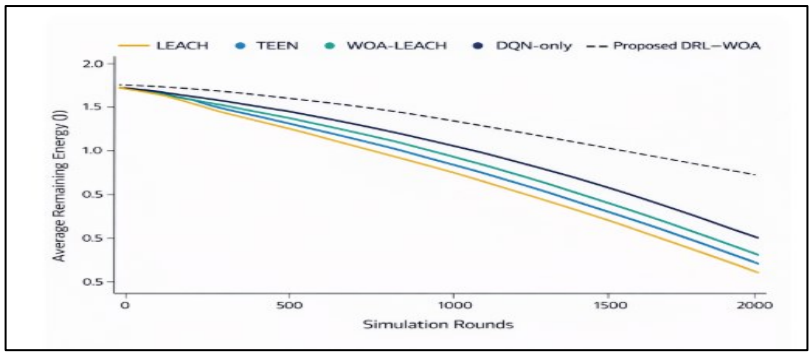

Fig. 3 : *Energy Consumption*

- **Packet Delivery Ratio PDR**

The Packet Delivery Ratio PDR is an essential measure of how reliable data communication is within a Wireless Sensor Network WSN. PDR quantifies the quality of the routing decisions that are made at each time point in response to changing network conditions. In terms of the total packet delivery performance of each protocol during the entire course of the simulation, the results of the evaluation of the protocols are listed in table 6. From this table it can be seen that the proposed DRL-WOA framework provided the best PDR of all of the protocols evaluated including LEACH, TEEN, WOA-LEACH, and the DQN-only methods. The improved PDR is due to the end-to-end delivery of packets through the use of dynamic multi-hop routing decisions based on information learned by the DRL-agent. These decisions provide avoidance of congested and energy depleted path segments. A secondary method of assessing the time-dependent delivery consistency of each protocol was through the evaluation of the average per round PDR of each protocol. The average per round PDR is a useful metric because it provides insight into the number of times each protocol delivers packets successfully as well as the stability of the delivery performance of each protocol across simulation rounds. The proposed hybrid method demonstrated both higher packet delivery success rates than the other methods, as well as higher average per round PDR values (Table 7), providing evidence that not only does the proposed hybrid method deliver packets with higher success rates than the other methods, but it does so in a consistent manner throughout the simulations. Figure 4 illustrates the PDR trends of each of the protocols evaluated and clearly demonstrates the sustained advantages of the proposed hybrid method.

Table 6. *Packet Delivery Ratio PDR over the entire simulation duration*

Protocol	PDR (%))
LEACH	89.5
TEEN	91.1
WOA-LEACH	92.7
DQN-only	93.2
Proposed DRL-WOA	96.8

Table 7. *Comparison of overall and average per-round Packet Delivery Ratio PDR*

Protocol	Average PDR (%)
LEACH	78.3
TEEN	82.5
WOA-LEACH	87.2
DQN-only	88.5
Proposed Hybrid	93.4

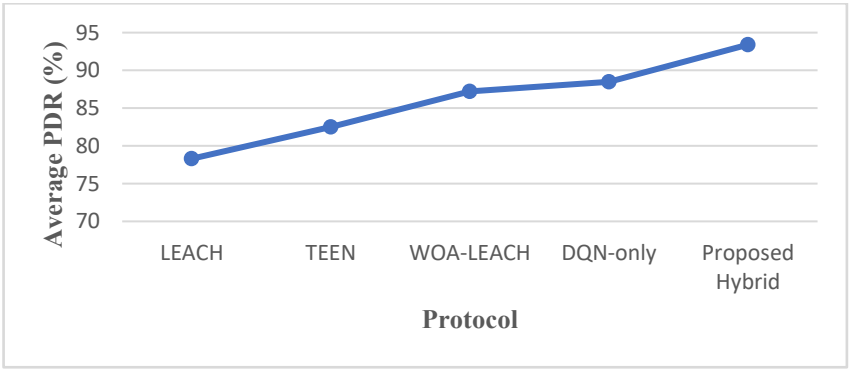

Fig. 4 *: Packet Delivery Ratio PDR*

- **Routing Overhead, Load Balancing, and DRL Agent Learning Time**

Routing Overhead is dependent on how well each of the routing schemes are able to adapt to changes and how many Control Messages that have to be sent (Figure 5). LEACH has the least amount of overhead because it uses a Static Probabilistic Routing Scheme. TEEN and WOA-LEACH have a moderate overhead for their Clustering and Coordinating functions. The DQN (only) will have the most overhead with regards to routing since there is always Route Exploration and Policy Updates occurring at

all times in the Learning Process. However, the Hybrid DRL – WOA Framework is able to maintain a controlled amount of routing overhead by finding a Balance between the adaptability that comes from learning and the control message overhead. Maintaining this type of balance allows for Improved Load Distribution and Stable Network Operation without having to incur too much overhead in terms of Communication.

In this work, a slight increase in overhead was observed due to the exploration phase of the DRL agent and the control messages required for joint coordination between routing and aggregation. While this overhead slightly rises control traffic, the trade-off is justified by the gains in adaptability, energy efficiency, and network lifespan. This study faces some issues (limitations), such as assuming fixed sensor nodes, relying on computationally intensive training carried out off-system, and not having been validated on very large scales. Future work will address these issues.

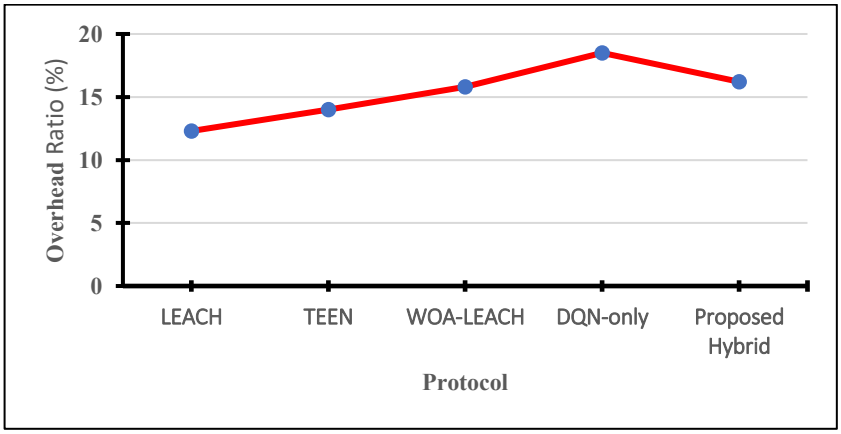

Fig. 5 : *Routing Overhead Ratio*

Performance of load balancing is measured by comparing the amount of energy that is consumed by the cluster heads at different times. In this way, the performance of the proposed DRL-WOA method is illustrated in Fig. 6, where the variance of the residual energy levels of cluster heads over time are compared to those of the other methods. The DRL-WOA method demonstrates a greater uniformity in energy consumption than the other methods because of the ability of the bio-inspired WOA optimizer to perform global searches. By having all cluster heads consume an equal or near-equal amount of energy, the DRL-WOA method can prevent certain nodes from being depleted of their energy resources at a faster rate than others and increase the overall longevity of the network.

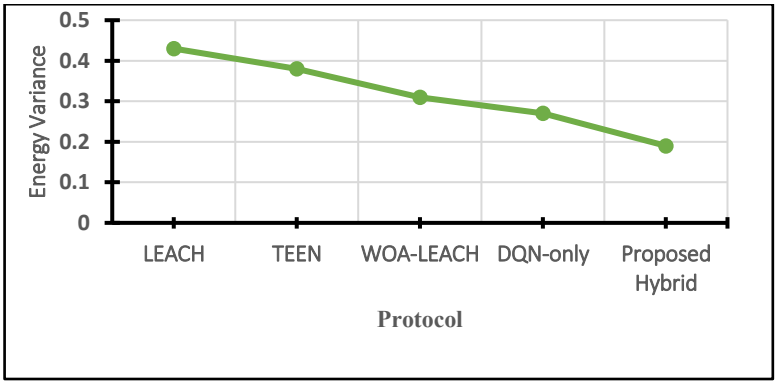

Fig. 6 : *CH Energy Variance*

Finally, the convergence properties of the DRL agent were studied to measure the quality of the learning process of the agent. Based on the results illustrated in fig. 7, the reward curve increases throughout the exploration period, but then begins to stabilize after the first 200 training periods. Therefore, the DRL agent was able to develop successful routing policies that provide better and more consistent energy efficiencies and routing decisions for the network over longer periods of time.

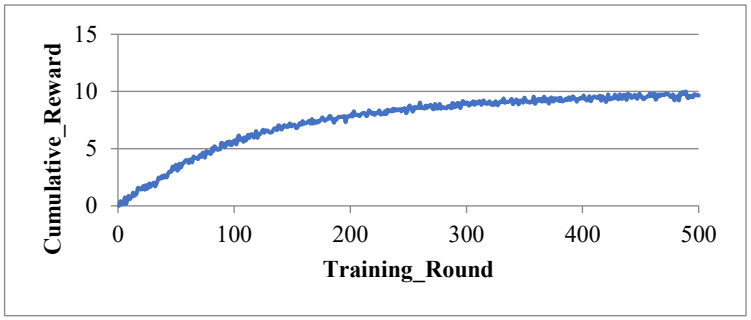

Fig. 7 : *DRL Agent Convergence*

6.1. Discussion

The results of the experiment show that the proposed DRL–WOA hybrid framework improves all the relevant metrics uniformly and optimally, such as energy saved, network lifetime, and the ratio of packets delivered. This model shows unyielding and flexible attributes regardless of the network density and varying conditions. The hybrid framework exceeds the performance of legacy routing and clustering by adapting to network configuration changes, and optimally shifting energy consumption to prolong the first node death, increasing network lifetime .

These results occur because the combination of WOA and DRL allows for a multi-tiered approach with the tier one being global cluster-heads where WOA finds the optimal cluster-heads by traversing the clusters to evaluate energy fitness, and tier two is policy discovery where DRL will learn routing policies in real-time based on the current state of the links and the energy status of each node. Optimizing both layers reduces redundant transmissions and the fluctuation in the cluster framework and improves packet delivery ratios under poor network conditions. However, the slight increase in routing overhead may indicate an opportunity for further optimizing the improvements of the system. Overall, the integrated DRL–WOA approach offers a more intelligent and scalable design for practical WSN applications, including smart agriculture, industrial monitoring, and disaster response, which prioritize energy efficiency and resilience over extended durations.

7. CONCLUSION

The hybrid DRL-WOA approach proposed here uses both DRL to optimize routing and WOA to select energy-efficient cluster heads to improve on prior methods that either do clustering / routing or only use one type of algorithm to perform both functions. Experimental results indicated that the framework presented here has better performance than LEACH, TEEN, WOA-LEACH, and a DQN-based only routing method regarding metrics such as energy consumption, packet delivery ratio, load distribution and network lifetime. The experimental results demonstrate that an integrated approach to the problems of energy aware clustering and routing can provide improvements over energy efficient and sustainable network operation in long term. Therefore, this proposed hybrid DRL-WOA approach is well-suited for WSNs operating under severe constraints in terms of available power where it is necessary to reliably operate in a dynamically changing environment. Despite promising results of the proposed method, the study has some limitations: namely, assuming that the sensor nodes are stationary, relying on computationally intensive offline training, and not being applied to very large scales. Future work will concern on expanding the framework to include scenarios that involve mobile nodes, including integrating node movement prediction into routing and aggregation decisions. In addition, exploring lighter DRL models is appropriate for on-device inference and validating the approach in real-world environments to assess scalability and practical deployment.

Conflicts of Interest

The author declares no conflicts of interest.

Funding

The author received no financial support for the research, authorship, and/or publication of this paper.

Acknowledgment

The author would like to thank Mustansiriyah University (www.uomustansiriyah.edu.iq) Baghdad-Iraq for its support in the present work.

REFERENCES

[1] G. Kumar, R. Lavanya, R. Monisha, and K. A., "An intelligent survey on usage of sensor network in real-time applications", Futuristic Trends in Network & Communication Technologies Volume 3 Book 2, IIP Series, pp. 335–349, 2024, doi: 10.58532/v3bgnc2p9ch1.

[2] A. Ramteke, "Wireless sensor network in environment monitoring: Advancements, applications, and challenges for real-time data collection and analysis," *Int. J. Multidisciplinary Research*, vol. 7, no. 3, 2025, doi: 10.36948/ijfmr.2025.v07i03.48803.

[3] D. K. Nishad and D. R. Tripathi, "Wireless sensor networks: Technologies and applications," *Turkish J. Comput. Math. Educ.*, vol. 11, no. 1, pp. 1673–1679, 2020, doi: 10.61841/turcomat.v11i1.14630.

[4] N. Vishnoi and R. K. Dwivedi, "Comprehensive survey on the reliability of wireless sensor networks," in Proc. SMART, pp. 496–502, 2022, doi: 10.1109/SMART55829.2022.10047428.

[5] S. K. Yadav, R. Sharma, and K. M. Adavala, "Toward sustainable wireless sensor networks: An integrated approach to energy conservation and protocol design," *Int. J. Sci. Adv. Technol.*, vol. 16, no. 3, 2025, doi: 10.71097/ijsat.v16.i3.7086.

[6] A.H.Abdul Kather, Dr. J. Karunanithi, "Energy optimization in wireless sensor networks: Trends, techniques, and trade-offs," *In book: AI in Industry 5.0: Revolutionizing Business and Technology*, pp. 93–98, 2025, doi: 10.26524/royal.239.19.

[7] S Sahu, S Silakari, "Energy efficiency and fault tolerance in wireless sensor networks: Analysis and review," *Soft Computing: Theories and Applications: Proceedings of SoCTA*, Springer, pp. 389–402, 2022, doi: 10.1007/978-981-19-0707-4_36.

[8] A. Ansari and B. Deshpande, "Enhancement of routing flexibility by a novel distributed approach for WSN," *IEEE 4th Annual Flagship India Council International Subsections Conference (INDISCON)*, pp. 1–8, 2023, doi: 10.1109/INDISCON58499.2023.10269878.

[9] M. Darbari et al., "An exhaustive review on advancements and challenges in low-power wireless sensor networks," Emerging Trends in Computer Science and Its Application, pp. 246–249, 2025, doi: 10.1201/9781003606635.

[10] P. C. Sridevi , R. S. Janaki, and G. Shanthini, "Future perspectives in energy-efficient wireless sensor networks: Exploring novel approaches," *Int. J. Adv. Netw. Appl.*, vol. 16, no. 4, pp. 6523–6532, 2025, doi: 10.35444/ijana.2025.16409.

[11] P. S. Prakash, D. Kavitha, and P. C. Reddy, "A rapidly-exploring random tree-based intelligent congestion control through alternate routing for WSNs," *Int. J. Commun. Netw. Distrib. Syst.*, vol. 29, no. 1, p.p. 71-94, 2023, doi:10.1504/IJCNDS.2023.127476.

[12] S. M. Tondare, V. Biradar, and M. M. Sardeshmukh, "Design and investigation of state-of-the-art WSN solutions for congestion control," *ShodhKosh: Journal of Visual and Performing*, vol.5, no. 6,p.p. 1661–1667, 2024, doi: 10.29121/shodhkosh.v5.i6.2024.2491.

[13] S. Basha *et al.*, "Congestion control routing scheme for WSN using AI technologies," in *Proc. ICEEICT*, pp. 1–6, 2024, doi: 10.1109/ICEEICT61591.2024.10718513.

[14] J. Thyagarajan, K. Suganthi, and G. Rajesh, "A joint congestion control mechanism through dynamic alternate route selection in IoT-based wireless sensor networks," Advances in Parallel Computing Algorithms, Tools and Paradigms,4I, p.p. 96- 102., 2022, doi: 10.3233/APC220013.

[15] P. Soltani, M. Eskandarpour, A. Ahmadizad, and H. Soleimani, "Energy-efficient routing algorithm for wireless sensor networks: A multi-agent reinforcement learning approach," arXiv preprint, arXiv:2508.14679, 2025, doi: 10.48550/arxiv.2508.14679.

[16] Z. Wang et al., "CRLM: A cooperative model based on reinforcement learning and metaheuristic algorithms of routing protocols in wireless sensor networks," Comput. Netw., vol. 236, 2023, doi: 10.1016/j.comnet.2023.110019.

[17] S. Sadhana, E. Sivaraman, and D. Daniel, "Enhanced energy-efficient routing using E-PEGASIS protocol for WSNs," Procedia Computer Science, vol. 194, pp. 89–101, 2022, doi: 10.1016/j.procs.2021.10.062.

[18] I. Daanoune, B. Abdennaceur, and A. Ballouk, "A comprehensive survey on LEACH-based clustering routing protocols in wireless sensor networks," Ad Hoc Netw., vol. 114, p. 102409, 2021, doi: 10.1016/j.adhoc.2020.102409.

[19] R. Ramya and K. Padmapriya, "Energy-efficient fuzzy-optimized routing in wireless sensor networks using PSO and WOA," J. Intell. Fuzzy Syst., vol. 44, no. 1, pp. 595–610, 2022, doi:10.3233/JIFS-220963.

[20] X. Xue et al., "A hybrid cross-layer Harris–Hawk-optimization-based efficient routing for wireless sensor networks," Symmetry, vol. 15, no. 2, p. 438, 2023, doi: 10.3390/sym15020438.

[21] M. Devika and S. M. Shaby, "Deep reinforcement learning-assisted butterfly optimization algorithm in MOD-LEACH routing for enhanced energy efficiency," Int. J. Comput. Eng. Sci. Eng., vol. 10, no. 4, pp. 1329-1336, 2024, doi: 10.22399/ijcesen.708.

[22] A. Nainwal et al., "Swarm intelligence-based clustering algorithms for wireless sensor networks," in Proc. IC3SE, Gautam Buddha Nagar, India, IEEE, pp. 1652–1657, 2024, doi: 10.1109/IC3SE62002.2024.10593090.

[23] V. Shobana. and J. Samraj, "Adaptive Reinforcement Learning-based Data Aggregation and Routing Optimization ARL-DARO for enhancing performance in WSNs," ICTACT J. Commun. Technol., vol. 15, no. 3, pp. 3282 – 3291, 2024, doi: 10.21917/ijct.2024.0488.

[24] V. Prakash, S. Pandey, and B. M. Sahoo, "Enhanced energy-efficient cluster-based routing with spotted hyena optimization in heterogeneous WSNs," EURASIP Journal on Wireless Communications and Networking., vol. 2025, no. 35, 2025, doi: 10.1186/s13638-025-02464-x.