SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I9P102

Volume 13 Issue 9, 12-22, September 2026

Original Article

A Hybrid Genetic Reinforcement Learning Technique for Intelligent Cluster Head Selection in WSNs

Ruchi Sharma[1], Sachin Patel[2]

[1,2]*Department of Computer Science and Engineering, Sage University, Madhya Pradesh, India.*

[1]*Corresponding Author : ruchi.sharma5aug@gmail.com*

Received: 25 May 2026 Revised: 06 August 2026 Accepted: 11 August 2026 Published: 29 September 2026

Abstract - *In Wireless Sensor Networks (WSNs), selection of Cluster Head (CH) in view of the optimum performance is a difficult task in view of constraints of energy resources, dynamic energy state of the network and dynamic data communication requirements, etc. Here, optimization of these can make better solutions. A Hybrid RL-GA Algorithm is developed to give the optimum solution for this, by combining a Genetic Algorithm (GA) with Q-learning. Optimization of both how CHs are selected and how the data is routed afterward was performed. Genetic Algorithm founds initial global structure, individual nodes and Cluster Heads take over using Q learning. This allows them to make adaptive routing decisions. Network continuously learns the most energy-efficient paths, and it recovers much faster from link failures. The GA provide a strong foundation, and Q-learning handles the day-to-day network stability. With simulations, our Hybrid RL-GA protocol outperformed the standalone RL method used in the latest research works. The total network lifetime is higher by roughly 50% than standalone RL.*

Keywords - *Wireless network, Cluster Head (CH) selection, Genetic Algorithm (GA), Reinforcement Learning (RL), Network lifetime.*

1. Introduction

The independent sensors scattered across a landscape, all working together to monitor parameters from sudden temperature drops to seismic tremors. These types of networks for sensors are termed Wireless Sensor Networks (WSNs). WSN is used for smart city infrastructure, modern healthcare monitoring, and even military reconnaissance.

However, there is a big issue of finite batteries of sensors deployed in remote or hard-to-reach places. [1, 2]. That makes energy-efficient data transmission the make-or-break factor for any WSN.

To handle this, engineers commonly use hierarchical clustering to manage the network's scale and power drain (see Figure 1) [3]. Instead of every node talking to the base station, the network breaks down into smaller groups.

Standard member nodes send their data to a local Cluster Head (CH). The CH then gathers this raw data and acts as the messenger, forwarding it to the Base Station (BS), which eventually connects everything to the internet.

The challenge of the optimization is mainly the dynamics of the network. As time passes, many nodes of the network become dead, and that affects the data transmission, the number of nodes in a cluster, energy losses, etc.

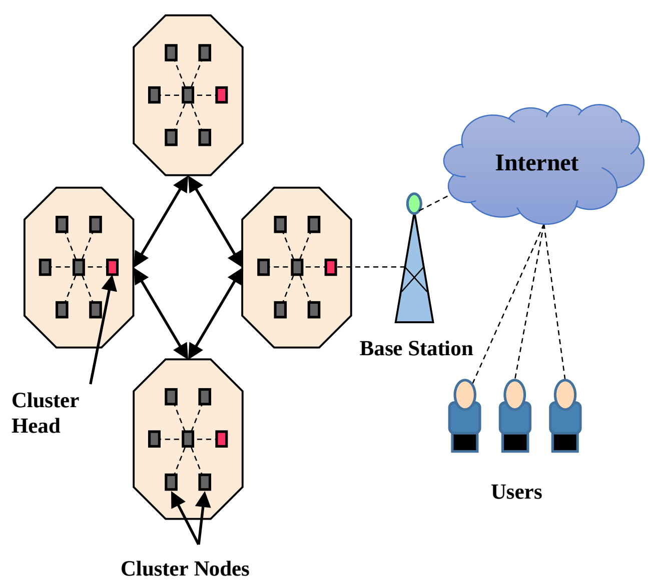


Fig. 1 Typical configuration of WSN

1.1. Problem Statement and Research Gap

Because sending and receiving data is the biggest energy drain for a sensor, how these Cluster Heads are chosen in this dynamic situation directly determines how long the network survives. If the selection process is flawed, certain nodes deplete their batteries too rapidly, creating the "hotspot problem" where dead zones fragment the network. While Reinforcement Learning (RL) has been utilized to optimize routing decisions on the fly, standalone RL often suffers from a lack of global perspective, easily falling into local optima during cluster formation. Conversely, pure metaheuristic approaches like Genetic Algorithms are computationally heavy and lack real-time adaptability to sudden network topology changes.

1.2. Novelty and Contributions

To get the best of both worlds, this study pairs the broad search power of a Genetic Algorithm (GA) with the real-time adaptability of Q-learning. Critical data, such as a node's remaining battery, local node density, and distance to the BS, is fed into both the GA's fitness function and the RL's reward system. This approach balances energy loss with efficient data movement.

The algorithms support each other: the GA provides the RL agent with a highly optimized global starting point, and the RL agent takes over to manage dynamic, day-to-day routing changes. This manuscript expands upon foundational standalone RL concepts previously presented by the researcher, introducing the novel integration of the GA phase, comprehensive algorithmic formulations, and extended complexity metrics.

2. Related Work

The paper deals with advanced techniques of AI for providing a solution to a well-known issue of the WSN. This section of the paper provides insights into the problem statement, followed by the techniques used in this paper and the literature about a similar approach adapted by other researchers.

2.1. WSN Issue: Foundational Energy Management and Optimum Routing Protocols

The finite battery life of nodes is still a major constraint for WSN, hence protocols are required for minimizing energy dissipation during communication and data aggregation [1, 3]. The initial important effort was for establishing a classification of routing techniques, sorting them into data-centric, hierarchical, and location-based approaches [2]. The Low-Energy Adaptive Clustering Hierarchy (LEACH) protocol turned out to be a point of reference by utilizing a randomized alternation of Cluster Heads (CHs) to distribute the energy load [4]. Subsequent revisions for the protocols were invented for advanced CH election processes, highlighting that rotation occurrence must be dynamically used to prevent fast depletion in high-density regions [5, 6, 7]. Recent reviews emphasize that while these protocols are initial, they do not support the scalability required for modern, high-density Internet of Things (IoT) applications [8]. There is also research for optimizing the communication structure through gateway nodes [9].

2.2. Meta-Heuristic Optimization via Genetic Algorithms

In WSN, optimization of node deployment along with energy consumption is an NP-hard problem; to solve this, researchers have gradually turned to Genetic Algorithms (GAs) [10, 11]. GAs allow for a comprehensive search of the solution space for identifying the global maxima or minima. It is established from research papers that GA-inspired selection methods significantly outperform randomized protocols by evaluating multi-objective fitness functions—balancing residual energy, node density, and distance to the sink [12, 13]. In 2026, advancements in GA were introduced with Two-Phase Distributed Genetic-based Algorithms (2PDGA) to solve scalability and deployment limitations in industrial sensor networks, moving beyond simple routing to consider hardware-specific constraints and time-sensitive scheduling [13]. The swarm intelligence is also a promising technique for optimizing the performance of WSN [14].

2.3. Transition to Artificial Intelligence and Neural Networks

WSN environments have a dynamic structure, this shift the research focus toward Artificial Intelligence (AI) to enable self-governing decision-making [15]. Initially, integrations of AI focused on using Knowledge-based systems for predicting the link quality [16]. This process evolved into the use of Artificial Neural Networks (ANNs) for energy /power management and for cluster selection [17, 18]. Modern research has extended this into 6G-enabled networks, where AI is used for real-time traffic prediction and energy harvesting optimization from ambient sources such as solar and Radio Frequency (RF) [19, 20]. This approach has improved the identifying non-linear patterns in energy consumption [18, 21].

2.4. Reinforcement Learning and the Density-Awareness Gap

Further, there is a need to achieve a balance between computational efficiency and adaptability. In this case, Reinforcement Learning (RL) has become the modern standard for WSN routing [22-24]. Using Q-learning, nodes can autonomously learn energy-efficient paths by environmental interaction [23]. However, a continuous challenge in RL models is their tendency to overgeneralize the situation by focusing exclusively on local distance [25]. Protocols like ReLeC [25] and RLBEEP [26] have introduced clustering-enhanced rewards to handle this issue. RL frameworks fail to account for how local node density interacts with global network connectivity-a gap our unified RL framework specifically addresses [26, 27].

2.5. Emerging Frontiers: 6G and Hybrid Intelligence

Presently, the research has shifted toward "intelligence native" 6G networks, which require ultra-low latency and massive connectivity [20, 27]. Recent breakthroughs include the FMADRL system, a scalable and QoS-aware reinforcement-learning framework. This framework is designed for highly dynamic mobile WSNs [19, 28]. Furthermore, hybrid meta-heuristic algorithms - such as the Hybrid Gazelle Optimization and Reptile Search Algorithm (HGORSA) - have achieved improvements in stability periods of up to 55% compared to standalone methods [29].

Selecting the ideal subset of nodes to act as Cluster Heads (CHs) is an NP-hard problem. Traditional probabilistic methods like LEACH often cause early node failure because they do not account for residual energy during the selection process. To solve this, recent 2026 research focuses on advanced hybrid meta-heuristics. Dhaka et al. (2026) developed an Adaptive Hybrid GWO-PSO framework that balances global and local searches using energy, position, and load metrics [30].

Similarly, Mahanta et al. (2026) introduced a Quantum-inspired Hybrid (QPSO-GA) framework. This system uses multi-objective fitness functions to structure clusters based on node density and transmission distance [31]. To improve search diversity, the Enhanced Secretary Bird Optimization Algorithm (ESBOA) was introduced in 2026 [32], utilizing chaotic maps and node degrees to prolong the network's lifetime. Additionally, the Hybrid Firefly-Bat Algorithm (HFBA) was proposed in 2026 [33], combining fast tracking with broad searching to lower cumulative energy consumption by 23.6%.

Because static heuristic models struggle with unpredictable topological variations in volatile WSN spaces, recent research has embedded Machine Learning (ML) directly into sensor nodes for real-time Cluster Head (CH) deployment. Modern frameworks utilize Multilayer Perceptrons (MLPs) and deep neural architectures to compute optimal CH coordinates by evaluating changing environmental factors and residual power levels.

Maddikera Krishna Reddy et al. introduced a dual-layer approach combining Chaotic Particle Swarm Optimization (CPSO) for initial CH allocation with an Attention-based Recurrent Neural Network (RNN) to manage downstream data routing [34]. This architectural division allows meta-heuristics to handle spatial clustering organization while deep learning models map complex, non-linear data transmission patterns.

In recent research, Reinforcement Learning (RL) has become the benchmark for embedding autonomous adaptability within the WSN clustering layer. Unlike rigid heuristics, RL agents dynamically modify Cluster Head (CH) selection parameters by learning from cumulative environmental feedback.

In this direction, Harsh Vaghela et al. introduced a Proximal Policy Optimization (PPO) model within an OpenAI Gym setup for tracking long-term temporal energy depletion to select CH nodes using real-time density and communication boundaries [35].

2.6. Research Gap

The reported research works were done either by a meta-heuristic algorithm or by ML techniques. In this paper, the work is focused on building hybrid architectures by utilizing a unified RL framework [28] backed by a genetic algorithm for the optimization of the clustering process.

3. Methodology and Network Model

For the simulation work of the paper, a simulation featuring N identical sensor nodes scattered randomly across a 2D field (Xmax × Ymax meters), with a static Base Station (BS) anchored right in the middle, has been considered. For the decision-making process, local node density is directly embedded into the state space. Further, the final selection of a Cluster Head (CH) is determined. This is based on a composite score of learned Q-value and the node's proximity to the Base Station.

3.1. Network Configuration & Clustering

The Q-learning algorithm is used to decide whether the node becomes a cluster head or a regular node. Here, a significant upgrade by embedding the local node density directly into the state space is used for the decision. When forming clusters, nodes naturally try to link up with the CH that demands the least amount of communication energy. However, to prevent those cluster heads from dead immediately, we strictly enforce a fixed Cluster Size (Nc). This assurance the workload remains balanced across the entire grid.

3.2. Communication Hierarchy

Data moves through the network in two distinct stages: Member-to-CH (Local) and CH to BS (global). In the first case, regular nodes fed their data to their assigned CH with the energy cost ETx and ERx for transmission and reception. These energy costs here depend heavily on the simple Euclidean distance (d) between the sensor and its leader. In the second case, after gathering the local data, the CH has fed these data to the Base Station. A CH can transmit directly, but that burns a massive amount of power if the node is near the edge of the grid. To solve this, multi-hop relaying is allowed in the simulation. The data hops through other intermediary CHs to reach the center. While this spreads the energy load out nicely, it does mean we have to calculate the transmission and reception costs for every single hop along the chain.

3.3. Energy Consumption in a Round

For a Normal Node of the jth cluster, the energy can be given as:

$$\mathrm{Ei} = \mathrm{Es} + \mathrm{ETx}\left(\mathrm{k}, d_{n_i \rightarrow CH_j}\right) \tag{1}$$

Where, the notation $d_{ni \rightarrow CHj}$ represents the Euclidean distance from the normal node ni to its designated Cluster

Head CHj.

Es is the energy required for sensing. k is the size of the data packet (usually measured in bits).

In case the distance between the cluster head and its node is less than crossover distance d0 (free space model), then.

$$\mathrm{ETx}\ (\mathrm{k}, \mathrm{d}) = \mathrm{k}.\mathrm{Eel} + \mathrm{k}.\mathrm{ef}.\mathrm{d}^2 \tag{2}$$

Otherwise, it is a multi-path fading model and then

$$\mathrm{ETx}\ (\mathrm{k}, \mathrm{d}) = \mathrm{k}\ \mathrm{Eel} + \mathrm{k}.\mathrm{em}.\mathrm{d}^4 \tag{3}$$

Here,

Eel: Energy spent per bit by circuitry.

ef: Energy spent for a bit / m^2 for the free space model.

em: Energy spent for a bit/ m for the multi-path fading model.

For a CHj, energy consumption is given as

$$\mathrm{ECHj} = \mathrm{Es} + \sum_{\mathrm{i} \in \mathrm{CMs\ of\ j}} \mathrm{ERx}(\mathrm{k}) + \mathrm{Eproc}(\mathrm{NCMj} \cdot \mathrm{k}) + \mathrm{ETx}\left(\mathrm{kagg}, \mathrm{d}_{\mathrm{CH_j} \to \mathrm{BS}}\right) \tag{4}$$

Where ERx(k)=k·Eel;

For multi-hop

$$\mathrm{ECHj} = \mathrm{Es} + \sum_{\mathrm{i} \in \mathrm{CMs\ of\ j}} \mathrm{ERx}(\mathrm{k}) + \mathrm{Eproc}(\mathrm{NCMj} \cdot \mathrm{k}) + \mathrm{ETx\_path}(\mathrm{kagg}, \mathrm{path\ BS}) \tag{5}$$

Eproc for processing/aggregation might be a function of total received packets or total data. Node's Residual energy can be given as

$$E_{res}(\mathrm{t}) = \mathrm{E}_{\mathrm{initial}} - \sum \mathrm{E}_{\mathrm{consumed}}(\mathrm{t}) \tag{6}$$

This is a key point of this paper which needs to be maximize by reducing the energy loss. The energy loss depends on whether the node was a cluster head or a member and the distance.

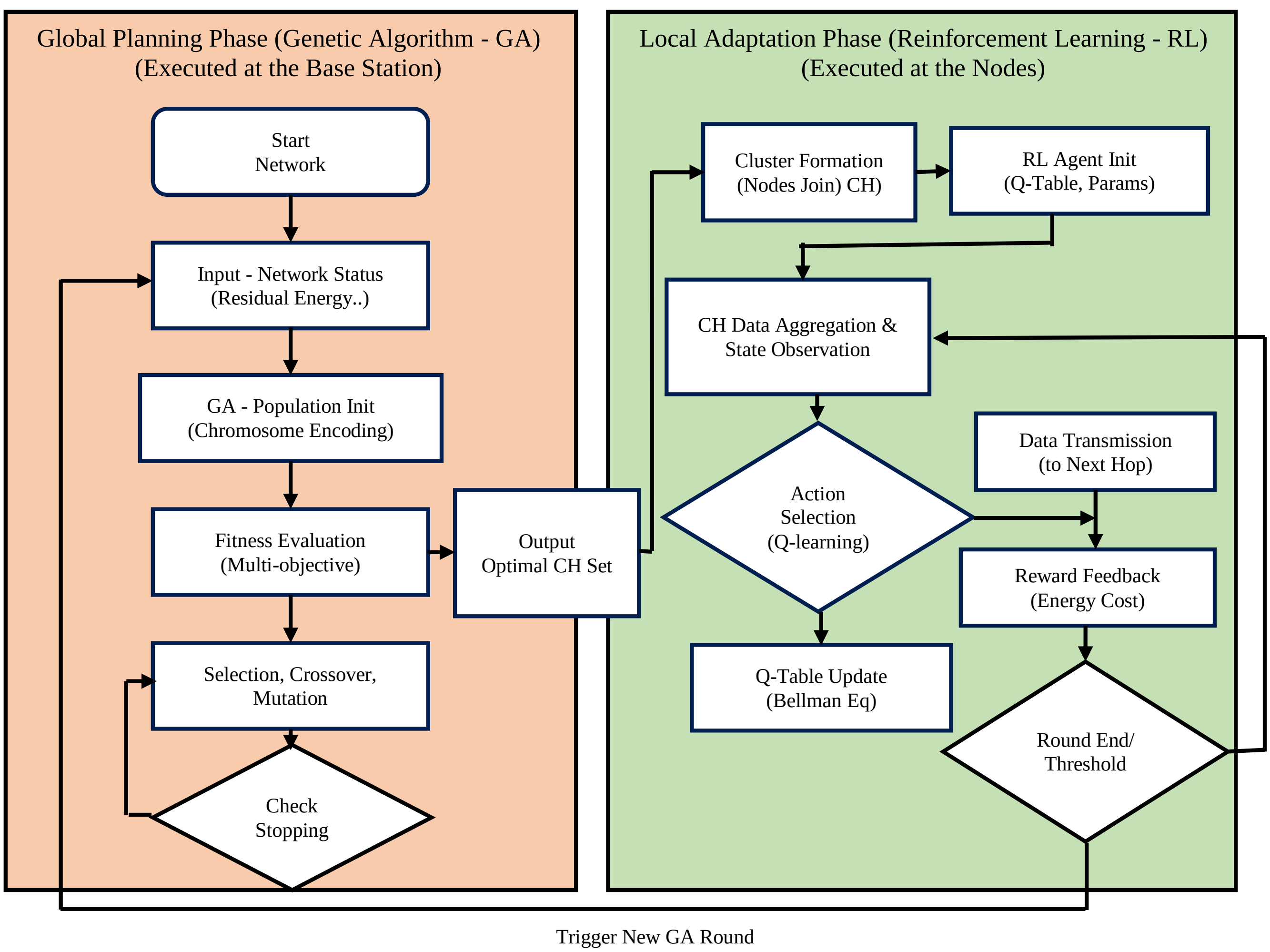


Fig. 2 Scheme of the efficient CH selection by the RL-GA framework

3.4. Simulation Framework

Reinforcement Learning (RL) and Genetic Algorithms (GA) based framework is typically divided into two distinct, cooperative phases: Global Optimization (Static/Global Planning) and Local Adaptation (Dynamic Routing) for the implementation of energy efficient Cluster Head (CH) selection method.

Rather than relying on a single metric, the BS evaluates these configurations using a multi-objective fitness function that balances three critical factors: Residual Energy (prioritizing nodes with high battery reserves), Density/Proximity (favoring centrally located CHs to reduce member communication costs), and CH-to-BS Distance (minimizing the transmission range for aggregated data).

3.4.1. Global Planning Phase (Genetic Algorithm)

The GA is executed at the Base Station to establish the initial global optimal CH selection. Chromosome Encoding: A binary string of length N represents the population, where '1' denotes a CH and '0' denotes a member node. Fitness Function: The fitness of each chromosome, which represents a candidate Cluster Head (CH) configuration, is evaluated using a multi-objective fitness function.

This function is designed to balance three primary network objectives: minimizing the total estimated energy consumption (E_{total}), maximizing the energy balance across the network by minimizing the standard deviation of residual energies (σ (E_{res})), and ensuring the selected number of Cluster Heads (N_{CH}) remains as close as possible to the optimal desired count ($N_{optimal}$). The overall fitness function F is mathematically defined as:

$$F = w_1\left(\frac{1}{E_{Total}+\varepsilon 1}\right) - w_2\ (\sigma\ \ (E_{res}).C1) + w_3 \max(0, 1 - \frac{|N_{CH}-N_{optimum}|}{N_{optimum}}).C2 \qquad (7)$$

Where: w1, w2, w3 denote the user-defined weights assigned to energy efficiency, energy balance, and connectivity, respectively. For this study, the weights are set to w1 = 0.4, w2 = 0.4, and w3 = 0.2. Epsilon ε1 is an infinitesimally small constant ($<10^{-9}$) introduced to prevent division by zero in the event of zero energy consumption.

C1 and C2 serve as scaling constants. These are required to mathematically normalize the standard deviation and connectivity ratio so that their numerical magnitudes align appropriately with the energy component prior to weight application. as:

For the GA application, a tournament selection operator is utilized to choose parent chromosomes from the population. Single-point crossover operator and a uniform mutation rate are applied to generate the new population.

3.4.2. Local Adaptation via Q-Learning

The GA base approach provides a strong structural foundation, whereas its static nature makes it slow to react to real-time disruptions like link failures or sudden energy drops. In our framework, it handles using Q-Learning at the node level.

State Space (S): The state represents the current node's environment, formulated as S = E_{res}, $Density_{local}$.

Action Space (A): A = {Become CH, Remain Member}.

Reward Function (R): The reward is evaluated such that nodes receive +10 for successful CH operation, +1 for member operation, scaled proportionally by energy conservation, and a strict -100 penalty upon node death.

Q-Update Equation:
$Q(s, a) = Q(s, a) + \alpha[R\ + \gamma\ max\ Q(s', a') -\ Q(s, a)]$ [8]

3.5. Hybrid Algorithm Workflow (Pseudocode)

To ensure full reproducibility of the proposed framework, the interaction between the global Genetic Algorithm optimization phase and the local Q-learning execution phase is detailed in Algorithm.

Algorithm: Hybrid RL-GA Protocol for CH Selection
Inputs: Network of N nodes, BS position, Initial Energy Einit, RL parameters (alpha, gamma, epsilon), GA parameters (P, G, w1, w2, w3).
Outputs: Optimal CH sets, Network Lifetime (FND, HND, LND), Cumulative Throughput.

Begin Simulation

For each independent random run (1 to 30) do:
  Initialize node positions, set E_{res}= E_{init} for all nodes.
  Initialize Q-table Q(S, A) = 0 for all States S and Actions A.
  For round = 1 to 5000 do:
   Phase I: State Observation & RL Action Preference
    For each alive node i do:
     Calculate the local density state based on the neighborhood radius.
     Determine energy state based on E_{res}.
     Form combined state S{t}= {Energy, Density}.
     With probability epsilon, select a random action At in CH, Member.
     Otherwise, select At = max Q(S{t}, a).
     If At == CH, add node i to Candidate CH List.
   End For
  Phase II: Global GA Optimization (Executed at BS)
  If Candidate CH List exceeds the desired optimal CH count, then:

Initialize Population P from Candidate CH List.
For generation = 1 to G do:
For each chromosome, do:
Evaluate Fitness F
End For
Select the best parents via Tournament Selection.
Apply Single-Point Crossover and Uniform Mutation.
Generate a new Population.
End For
Designate the best chromosome as the Actual CH Set.
Else:
Designate Candidate CH List as Actual CH Set.
End If
Phase III: Energy Dissipation and Q-Table Update
For each alive node i do:
If node i is in the Actual CH Set, then:
Calculate Econsumed for receiving, aggregating, and BS transmission.
Assign base reward R = +10.
Else:
Calculate Econsumed for local CH transmission.
Assign base reward R = +1.
End If
Update Eres = Eres - Econsumed .
If Eres < Threshold, apply the death penalty R = R - 100.
Otherwise, apply the energy conservation bonus to R.
Observe the next state S {t+1}.
Update Q-table:
End For
Decay exploration rate epsilon.
End For (Simulation Rounds)
End For (Independent Runs)
End Simulation

4. Simulation and Results

The proposed methodology is simulated using MATLAB. The key simulation parameters are:

- Network Size: 50 nodes
- Network Area: 100m x 100m
- Initial Node Energy: 0.5 Joules
- Minimum Energy Threshold: 0.01 Joules
- Base Station Position: [50, 50] (center of the network)
- Data Packet Size: 4000 bits
- Energy States: 5
- Density States: 5 (with Radius of 25m)
- Learning Rate (α): 0.1
- Epsilon Decay: 0.995
- Minimum Epsilon: 0.1
- Initial Epsilon (ϵ): 0.9
- Discount Factor (γ): 0.9
- Simulation Rounds: 5000
- Desired CHs: Max 10% of alive nodes (min 1 CH)

A simulation with the above parameters was done, including RL-GA-selected CHs for each round separately. The no. of CHs, total alive nodes, total dead nodes, and remaining energy were recorded. The spatial distribution of the nodes with the BS is shown in Figure 3.

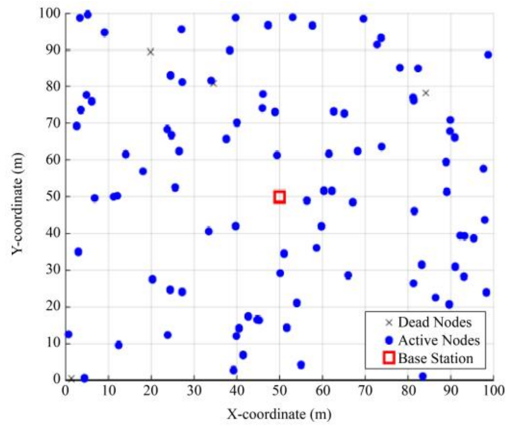
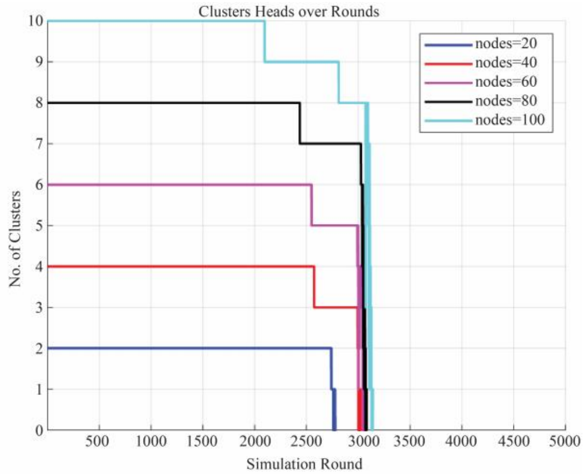


Fig. 3 Typical distribution of nodes (active and dead nodes) and BS during simulation

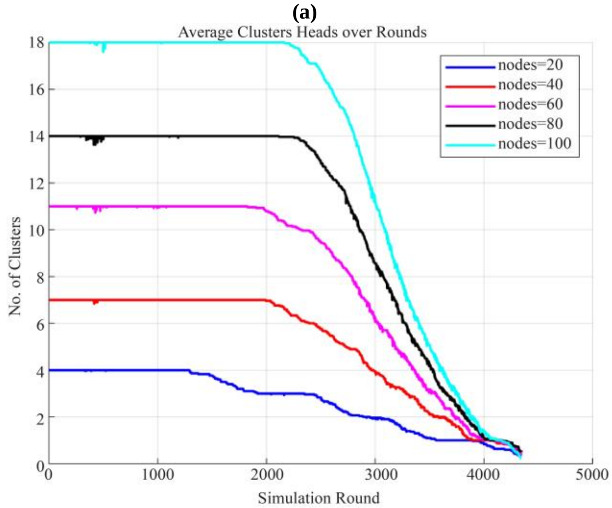
4.1. Performance Analysis Compare to the Standalone RL Method

Figure 4 shows the number of clusters formed over rounds in the case of different no. of nodes for stated simulation parameters.

Figure 4 presents a comparative analysis of the number of active Cluster Heads (CHs) over the course of simulation rounds for varying network densities (20, 40, 60, 80, and 100 nodes). The results contrast the performance of the standalone Reinforcement Learning (RL) method (Figure 4(a)) against the proposed hybrid Reinforcement Learning and Genetic Algorithm (RL-GA) approach (Figure 4(b)).

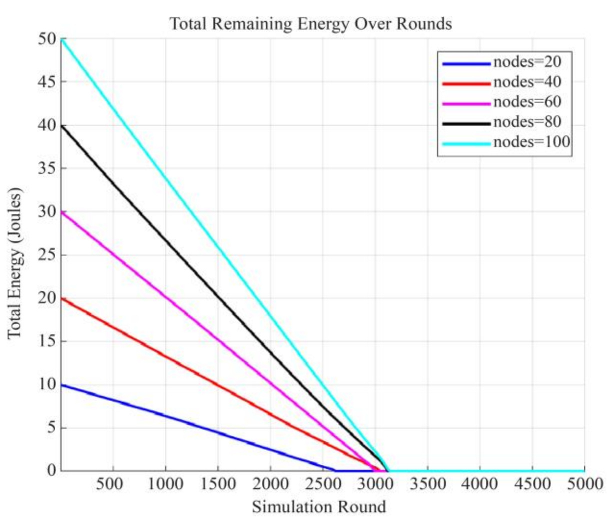


(a)

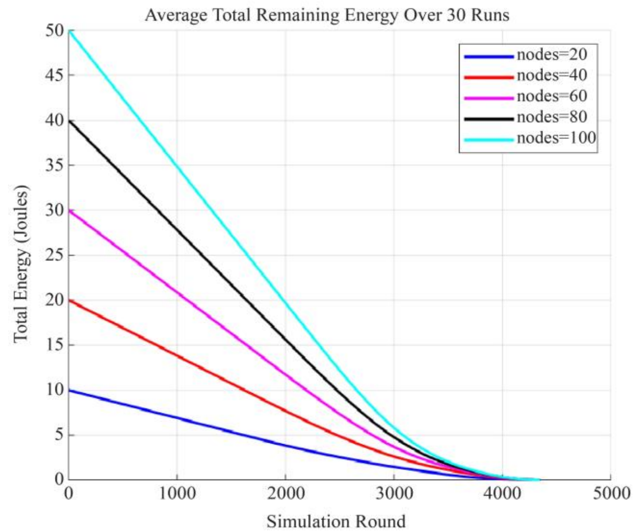


(b)

Fig. 4 Typical graph for the number of clusters over rounds for different no. of nodes (a) RL [28], and (b) RL-GA.



(a)



(b)

Fig. 5 Total remaining energy over simulation rounds for, (a) RL [28], and (b) RL-GA.

Figure 5 presents the variation in the network's total residual energy over the simulation rounds for varying node densities (20, 40, 60, 80, and 100 nodes).

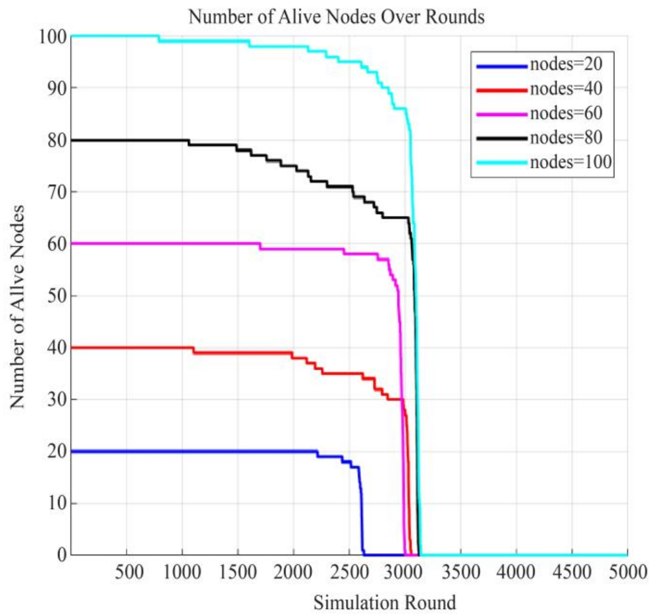
This metric is critical for evaluating the energy efficiency of the routing protocols, comparing the standalone Reinforcement Learning (RL) method (Figure 5(a)) against the proposed Hybrid RL-GA approach (Figure 5(b)).

Energy Depletion in Standalone RL (Figure 5(a)): In the RL-based scenario, the total energy consumption follows a near-linear downward trend. The steep, constant slope indicates that the energy consumption rate remains high and unmitigated throughout the network's life.

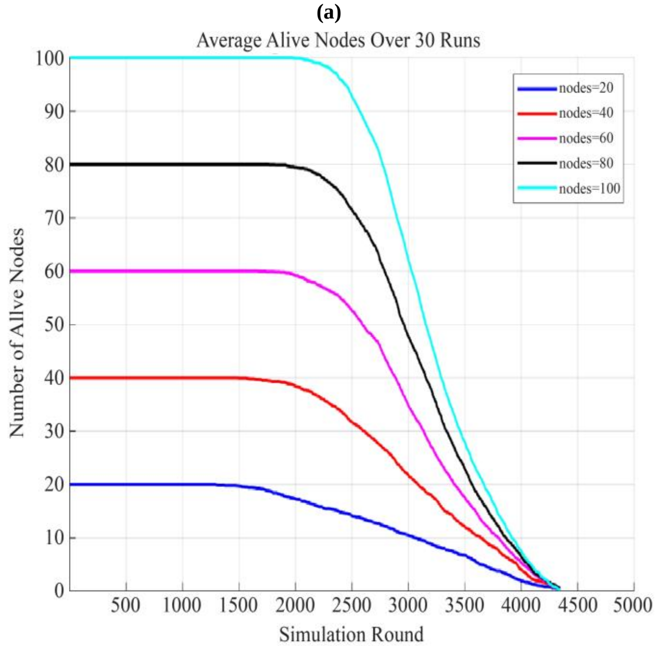
For the high-density scenario (100 nodes, cyan line), the total initial energy of 50 Joules is completely exhausted by approximately 3,200. This rapid, linear decay suggests that the RL algorithm lacks the dynamic optimization capabilities required to conserve energy during critical low-energy phases.

Energy Conservation in Hybrid RL-GA (Figure 5(b)): The Hybrid RL-GA method exhibits a distinctly different energy consumption profile. While the initial phase shows similar consumption to the RL method, the curve becomes less steep as the simulation progresses, particularly in the latter half (after round 2,500).

This "flattening" of the curve near the end-of-life stage indicates that the Genetic Algorithm is successfully optimizing cluster head selection to preserve the remaining energy. Consequently, the network lifetime for 100 nodes is extended significantly, reaching approximately 4,200 rounds.



(a)



(b)

Fig. 6 Number of alive nodes over rounds for, (a) RL [28], and (b) RL- GA.

Figure 6 presents the network survivability profile by plotting the number of alive nodes against simulation rounds for various network densities (20 to 100 nodes). This metric is essential for determining the "First Node Dead" (FND) and "Last Node Dead" (LND) operational points, comparing the standalone RL method (Figure 6(a)) to the Hybrid RL-GA method (Figure 6(b)).

Sudden Failure in Standalone RL (Figure 6(a)) The RL-based protocol demonstrates a "rectangular" survival curve, characterized by high stability followed by catastrophic failure. For the 100-node scenario (cyan line), the network retains nearly 100% of its nodes until approximately round 2,800.

However, once failures begin, the curve exhibits a near-vertical drop, with the entire network collapsing (LND) by round 3,200. This "sudden death" phenomenon implies that the RL algorithm depletes the energy of all nodes at a nearly identical rate. While this indicates load balance, it results in a total cessation of network coverage abruptly, leaving no residual sensing capability.

Extended Lifetime via Graceful Degradation in RL-GA (Figure 6(b)): The RL-GA approach demonstrates a significantly different mortality curve, characterized by "graceful degradation."

Although the onset of node failure (FND) occurs slightly earlier than in the RL method (around round 2,200 for 100 nodes), the slope of the decline is much shallower. The Genetic Algorithm's optimization prevents the simultaneous exhaustion of all nodes, allowing stronger nodes to compensate for weaker ones. Consequently, the network remains operational for a much longer duration, with the (LND) approximately 4,300 rounds.

The comparative analysis with other similar work done to improve the performance of WSN in terms of energy of the network is summarized in the Table 1.

Table 1. Comparative results for WSN lifetime with respect to the proposed method (average over 30 simulation run)

Protocol / Algorithm	First Node Dead (FND)	Half Node Dead (HND)	Last Node Dead (LND)
Hybrid RL-GA (Student)	~2,200 rounds	~3,200 rounds	~4,300 rounds
Standalone RL / RLBEEP [26].	~2,800 rounds	~3,000 rounds	~3,200 rounds
EAC-ECHS (Grid) [6].	~1,800 rounds	~2,400 rounds	~4,000 rounds
ANN-based Selection [21].	~200 rounds	~1500 rounds	~4,000 rounds

4.1.1. Computational Complexity and Parameter Sensitivity

Computational Complexity

The standalone Q-learning algorithm processes in O(N) time per round. The introduction of the GA at the base station adds a complexity of O (P . G . N), where P is the population size, and G is the number of generations.

Because the GA is centralized and only triggered at specific energy thresholds, the daily operational complexity at the node level remains O(1) for state lookups, ensuring high efficiency.

Parameter Sensitivity

Sensitivity analysis revealed that maintaining a low learning rate (alpha = 0.1) prevents the model from overriding the GA's structural foundation too rapidly. Furthermore, varying the GA mutation rate demonstrated that values between 0.01 and 0.05 optimally balance exploration without destroying high-fitness topologies.

4.2. Detailed Discussion of Results

The reported improvements over state-of-the-art benchmarks stem directly from mitigating the RL agent's local optima trap. Standalone implementations like RLBEEP show excellent early stability but result in massive simultaneous failures. By periodically injecting a globally optimal topology via the GA, the proposed framework forces the Q-learning agents to redistribute communication loads across underutilized sectors of the network, fundamentally resulting in the observed extended HND and LND metrics. The results for different nodes are shown in Table 2.

Table 2. Comparative results for different nodes (average over 30 simulation run)

Nodes	Avg FND	Avg HND	Avg LND	Avg Throughput (Packets)
20	1604	2997	4250	11756
40	1820	3071	4332	22305
60	1972	3120	4320	34757
80	2037	3147	4345	45157
100	2170	3153	4298	57789

From Table 2, it is clearly observe a linear throughput with the number of nodes and a nearly similar lifetime. This indicates the optimum use of energy. The FND and HND are also increased with the number of nodes.

5. Conclusion

This paper presents an innovative Hybrid Reinforcement Learning and Genetic Algorithm (RL-GA) framework aimed to improve energy efficiency and prolong the operational lifetime of WSN. The proposed technique addresses the limitations of standalone Reinforcement Learning (RL) by integrating the global search and optimization capabilities of Genetic Algorithms (GA) to dynamically manage CH selection.

The simulation results conclusively demonstrate the superiority of the Hybrid RL-GA protocol over the standalone RL method across varying network densities. The analysis of cluster stability (Figure 4) revealed that while the RL method relies on a static and rigid clustering policy, the Hybrid RL-GA approach employs a dynamic, step-wise reduction in cluster counts. This adaptive behavior allows the network to optimize the load distribution in response to the depleting energy reserves of the nodes. Furthermore, the energy depletion analysis (Figure 5) and node survivability profiles (Figure 6) confirm that the hybrid approach significantly mitigates the "sudden death" phenomenon observed in traditional RL methods. By promoting graceful degradation, the Hybrid RL-GA algorithm successfully extends the network lifetime (Last Node Dead) from approximately 3,200 rounds to 4,300 rounds in high-density scenarios—an improvement of nearly 35%. Overall, the hybridization of RL and GA provides a robust, energy-efficient routing solution that maximizes the utility of limited network resources.

6. Study Limitations and Future Study

Despite the several advantages of the proposed GA-RL paradigm, there are some architectural bottlenecks. One notable bottleneck is the computational scalability constraints for executing GA at the base station in the case of dense topologies. The other one is the transient packet delivery drops during the dispersed Q-learning "cold start" stage. These operational vulnerabilities are amplified for the rigid, threshold-driven re-clustering schedule.

In future, these limitations can be resolve using upgrading the tabular RL core to deep neural architectures (such as DRL). This would allow the framework to capture continuous, non-linear environmental states. Furthermore, improving the approach with security by measuring how secure and trustworthy the devices are, along with testing it on real physical IoT devices, will be a key next step.

Conflicts of Interest

“The author affirms that there is no conflict of interest regarding the publication of this paper.”

Funding Statement

“This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.”

References

[1] Mohd Fauzi Othman, and Khairunnisa Shazali, “Wireless Sensor Network Applications: A Study in Environment Monitoring System,” *Procedia Engineering*, vol. 41, pp. 1204-1210, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[2] J.N. Al-Karaki, and A.E. Kamal, “Routing Techniques in Wireless Sensor Networks a Survey,” *IEEE Wireless Communications*, vol. 11, no. 6, pp. 6-28, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[3] Vivek Mhatre, and Catherine Rosenberg, “Design Guidelines for Wireless Sensor Networks: Communication, Clustering and Aggregation,” *Ad Hoc Networks*, vol. 2, no. 1, pp. 45-63, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[4] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, Maui, HI, USA, vol. 2, pp. 1-10, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[5] Jun Wang et al., "A Cluster-Head Rotating Election Routing Protocol for Energy Consumption Optimization in Wireless Sensor Networks," *Complexity*, vol. 2020, no. 1, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] El Idrissi Nezha, Najid Abdellah, and El Alami Hassan, "Energy-Aware Clustering and Efficient Cluster Head Selection," *International Journal on Smart Sensing and Intelligent Systems*, vol. 14, no. 1, pp. 1-15, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ahmed A. Al-Healy, and Qutaiba Ibrahim, "WSN Routing Protocols: A Clear and Comprehensive Review," *International Journal of Advanced Natural Sciences and Engineering Researches*, vol. 9, no. 3, pp. 1-14, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[8] Rajesh Mitukula, Shashank Devasani, and Manasa Bolumalla, "Scalable and Energy Efficient Routing Protocol for Wireless Sensor Networks," *Engineering Research Express*, vol. 8, no. 1, pp. 1-10, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[9] G.H. Raghunandan, S. Vishali, and R. Raman, "Design and Development of an Efficient Routing Protocol for WSN," *2025 5th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, Mandya, India, pp. 1-6, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Yousra Ben Aissa, Wafa Bouaziz, and Amira Mohammedi, "Genetic Algorithm-based Optimal Wireless Sensor Network Deployment," *2022 International Symposium on iNnovative Informatics of Biskra (ISNIB)*, Biskra, Algeria, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ali Norouzi, and A. Halim Zaim, "Genetic Algorithm Application in Optimization of Wireless Sensor Networks," *The Scientific World Journal*, vol. 2014, no. 1, pp. 1-15, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[12] Biswa Mohan Sahoo, Hari Mohan Pandey, and Tarachand Amgoth, "A Genetic Algorithm Inspired Optimized Cluster Head Selection Method in Wireless Sensor Networks," *Swarm and Evolutionary Computation*, vol. 75, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Bartłomiej Płaczek, and Marcin Bernaś, *Architectures, Protocols and Algorithms of Sensor Networks*, 2nd ed., Sensors, vol. 26, pp. 1-250, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[14] K. Ananthi et al., "Swarm Intelligence-based Clustering For Self-Healing Wireless Sensor Networks," *International Journal of Advances in Signal and Image Sciences*, vol. 12, no. 1, pp. 1-13, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[15] Qianao Ding et al., "An Overview of Machine Learning-Based Energy-Efficient Routing Algorithms in Wireless Sensor Networks," *Electronics*, vol. 10, no. 13, pp. 1-24, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Julio Barbancho et al., "Using Artificial Intelligence in Wireless Sensor Routing Protocols," *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, vol. 4251, pp. 475-482, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[17] Ahmad Hosseingholizadeh, and Abdolreza Abhari, "A Neural Network Approach for Wireless Sensor Network Power Management," *28th IEEE International Symposium on Reliable Distributed Systems*, Niagara Falls, NY, USA, pp. 301-302, 2009. [Google Scholar] [Publisher Link]

[18] Neda Enam et al., "Neural Network Based Energy Efficiency in Wireless Sensor Networks: A Survey," *International Journal of Computer Science and Engineering Survey*, vol. 1, no. 1, pp. 39-53, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[19] Rakan Alanazi et al., "Machine Learning-Driven Routing Optimization for Energy-Efficient 6G-Enabled Wireless Sensor Networks," *Alexandria Engineering Journal*, vol. 129, pp. 877-888, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[20] Muhammad Owais, and Thokozani Shongwe, "Machine Learning-Enabled 5G and 6G Networks: Methods, Challenges, and Opportunities," *Applied Sciences*, vol. 16, no. 4, pp. 1-40, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[21] Volkan Gangal, Erhan Sesli, and Gokce Hacioglu, "Artificial Neural Network based Cluster Head Selection Process in Wireless Sensor Networks," *Cluster Computing*, vol. 29, pp. 1-20, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[22] Anju Arya, and Nisha Nehra, "Reinforcement Learning based Routing Protocols in WSNs: A Survey," *International Journal for Research in Applied Science and Engineering Technology*, vol. 6, no. 4, pp. 1-9, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[23] Zhang Zhaohui, Zhou Jiaqi, and Li Jing, "Q-Learning-based Semi-Fixed Clustering Routing Algorithm in WSNs," *Ad Hoc Networks*, vol. 174, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[24] Yu Song et al., "Research on High-Efficiency Routing Protocols for HWSNs Based on Deep Reinforcement Learning," *Electronics*, vol. 13, no. 23, pp. 1-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Tripti Sharma et al., "ReLeC: A Reinforcement Learning-Based Clustering-Enhanced Protocol for Efficient Energy Optimization in Wireless Sensor Networks," *Wireless Communication and Mobile Computing*, vol. 2022, no. 1, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] Ali Forghani Elah Abadi et al., "RLBEEP: Reinforcement-Learning-Based Energy Efficient Control and Routing Protocol for Wireless Sensor Networks," *IEEE Access*, vol. 10, pp. 44123-44135, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[27] Marcin Lewandowski, and Bartłomiej Płaczek, "A Cluster Head Selection Algorithm for Extending Last Node Lifetime in Wireless Sensor Networks," *Sensors*, vol. 25, no. 11, pp. 1-31, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[28] Ruchi Sharma, and Sachin Patel, "Reinforcement Learning Strategies for Energy Efficient Cluster Head Selection in WSNS," *Adroid Conference Series: Engineering and Technology*, vol. 1, no. 1, pp. 66-76, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[29] Rakesh Kumar Yadav, and Rajendra Prasad Mahapatra, "Hybrid Metaheuristic Algorithm for Optimal Cluster Head Selection in Wireless Sensor Network," *Pervasive and Mobile Computing*, vol. 79, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[30] Raghuveer Singh Dhaka et al., "Energy–Lifetime Aware Cluster Head Selection in Wireless Sensor Networks using an Adaptive Hybrid GWO–PSO Framework," *Discover Computing*, vol. 29, pp. 1-28, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[31] Ayush Mahanta et al., "Quantum-Inspired Hybrid Optimization Framework for Energy-Efficient Clustering and Routing in Wireless Sensor Networks," *Scientific Reports*, vol. 16, pp. 1-33, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[32] Amit Raj et al., "An Enhanced Secretary Bird Optimization Algorithm (ESBOA) for Global Optimization and its Applications to Engineering Design and Traveling Salesman Problems," *Journal of Engineering Research*, vol. 14, no. 2, pp. 1650-1682, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[33] J. Jegan et al., "Optimized Wireless Sensor Network Clustering Using a Hybrid Firefly–Bat Algorithm (HFBA)," *Journal of Computer Science*, vol. 22, pp. 1494-1509, 2026. [CrossRef] [Publisher Link]

[34] Maddikera Krishna Reddy et al., "Modeling of Metaheuristic-Based Dual Cluster Head Selection with Routing Protocol for Energy-Efficient Wireless Sensor Networks," *Soft Computing*, vol. 29, pp. 2999-3020, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[35] Harsh Vaghela et al., "Single Phase Photovoltaic Inverter with Grid Synchronization Using Dual Stage DC-DC Boost Converter," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 18, no. 15s, pp. 1816-1825, 2026. [CrossRef] [Publisher Link]