ADROID Conference Series: Engineering Sciences and Technology (ACS: EST)

International Conference on Artificial Intelligence for Sustainable Development, Jaypee University Anoopshahr, India 2025

Volume 01, October 2025, pp. 66-76

DOI: https://doi.org/10.63503/c.acset.2025.7

REINFORCEMENT LEARNING STRATEGIES FOR ENERGY EFFICIENT CLUSTER HEAD SELECTION IN WSNS

Ruchi Sharma[1] and Sachin Patel[2]

[1, 2] Department of Computer Science and Engineering, Sage University, Indore, India

ruchi.sharma@sageuniversity.in (https://orcid.org/0009-0000-8876-9516),
drsachin.patel@sageuniversity.in (https://orcid.org/0000-0003-0563-1752)

Abstract. Wireless Sensor Networks (WSNs) are important for numerous IoT applications. The limitation of the universal applications of WSN is due to the constrained energy reservoirs of the sensor nodes. For data communication, clustering of sensor nodes is an established scheme for minimizing the energy expenditure by appointing Cluster Heads (CHs) responsible for data aggregation and relay. This paper describes an approach for developing the cluster and CH selection dynamically for WSNs, leveraging Reinforcement Learning (RL), specifically Q-learning. The framework of this paper integrates residual energy level of a node and localized node density into the RL state space, empowering individual nodes to make perceptive decisions regarding their roles. Moreover, the CH selection mechanism prioritizes nodes that not only exhibit elevated Q-values for the CH function but are also optimum distance from the Base Station (BS), thereby minimizing cumulative transmission energy. The simulations show that this RL-driven, adaptive paradigm effectively equilibrates energy consumption throughout the network, resulting in an extended network lifespan and improved energy efficiency compared to traditional methodologies.

Keywords: Wireless Sensor Networks, Cluster Head Selection, Reinforcement Learning, Q-learning, Energy Efficiency, Network Lifetime, Node Density.

1 INTRODUCTION

WSNs, assemblages of spatially distributed, autonomous sensing devices, collaboratively monitor environmental parameters, such as thermometric fluctuations, acoustic discharges, vibrational oscillations, barometric gradients, kinematic displacements, and atmospheric contaminants. These networks have achieved widespread adoption across a wide range of domains, including ecological monitoring, industrial automation, healthcare, smart urban planning, and military investigation [1-2]. figure 1 shows a typical configuration of cluster based WSN. From figure 1, there are different clusters of WSN, in which nodes sent data to CH and these data are aggregated and communicate through CH. All CHs are connected to base station (BS). From BS, data are routed through the internet to end users. A critical challenge in WSNs is the constrained energy supply, which is frequently powered by depletable batteries and deployed in remote, inaccessible environments, thereby precluding the practicality of periodic battery replacement [3]. Reinforcement Learning (RL) presents a promising paradigm for addressing the dynamic and adaptive nature of CH selection. RL agents acquire optimal decision-making strategies by engaging in iterative trial-and-error interactions with their environment. This approach makes them well-suited for self-organizing and self-optimizing systems such as WSNs. This paper presents an RL-based framework for CH selection that integrates node energy, local density, and positional relevance to the BS to enhance energy efficiency and extend the overall network lifetime.

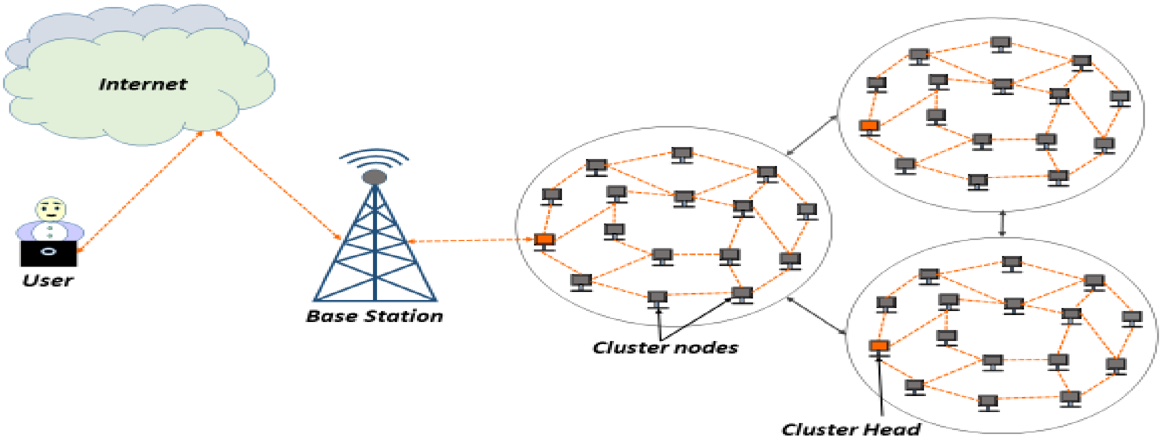

Figure 1. Typical configuration of WSN

2 RELATED WORK

The challenge of effective CH selection within WSNs has been subjected to rigorous academic investigation. Initial paradigms, exemplified by LEACH [4], propagated the innovative concept of rotational CH duty cycles to equitably disperse energy expenditure among nodes. Subsequent advancements, such as PEGASIS [5] and HEED [6], focused on enhancing energy conservation strategies through chain-oriented routing schemas or probabilistic CH election based on residual energy reserves and node connectivity degrees.

With the advancement of artificial intelligence paradigms, machine learning methodologies have progressively been harnessed to optimize WSN functionalities. Supervised learning techniques have been widely utilized for fault diagnostics and predictive analytics, while unsupervised techniques—particularly clustering algorithms, such as K-Means—have been scrutinized for optimal cluster allocation [7].

More recently, RL has gained traction due to its proficiency in managing dynamic environments and learn optimal policies without explicit programming. Q-learning, a model-free RL algorithm, has been applied to various WSN problems, including routing [8] and power control. For CH selection, some works have proposed Q-learning to choose CHs based on residual energy and distance [9]. However, many existing RL-based approaches for CH selection often simplify the network model or do not fully capture the multi-faceted nature of optimal CH selection, such as the combined impact of local density and global network topology. Our work aims to bridge this gap by integrating these critical parameters into a comprehensive RL framework.

3 PROPOSED METHODOLOGY

The proposed methodology employs a Q-learning algorithm to enable sensor nodes to intelligently decide whether to become a CH or remain a regular member node. The decision-making process is enhanced by incorporating node density into the state space and by refining the actual CH selection using a combined score that considers both the node's proximity to the BS and the learned Q-value.

3.1. Network Model

We conceptualize a WSN comprising N homogeneous sensor nodes randomly dispersed within a two-dimensional domain of dimensions Xmax × Ymax meters. A singular, static Base Station (BS) is situated at the centroid of this domain. Cluster affiliation is governed by an energy-aware protocol, whereby nodes join the Cluster Head (CH) that minimizes their communication energy expenditure. The Cluster Size (Nc), denoting the number of nodes within each cluster, remains fixed for this simulation to prevent disproportionate cluster sizes, which could induce load imbalance and cause Cluster Heads (CHs) to die out quickly. The communication paradigm for this WSN encompasses intra-cluster and inter-cluster transmissions. It includes:

Sensor Node to CH: Each standard Cluster Member Node (CM) sends its sensed data to the assigned CH. The energy consumption involved in this transmission is denoted as ETx for CMs, with ERx(k) representing the energy absorbed by the CH upon reception. The variable d indicates the Euclidean distance from the CM to its CH. Due to the proximal communication, energy expenditure is typically minimal.

CH to BS Transmission: Each CH transmits the aggregated data directly to the BS, requiring energy denoted as ETx1. This process can be highly energy-intensive, particularly for CHs positioned at greater distances from the BS.

Multi-hop Relay via CHs: CHs may relay data through intermediary CHs or dedicated relay nodes to reach the BS, summing the energy costs of transmission (ETx) and reception (ERx) across each hop. This routing mechanism employs protocols designed for CHs and often constitutes the most energetically demanding aspect of clustered WSN communication.

Energy Consumption per Round

For a Cluster Member (CM) in cluster j:

$$\mathrm{ECMi} = \mathrm{Esense} + \mathrm{ETx}\left(\mathrm{k}, d_{CM_i \to CH_j}\right) \tag{1}$$

Where, if d<d0: Crossover distance (free space model) then

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

if d>d0 : (multi-path fading model) then

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

Here,

Eelec: Energy consumed per bit by the transmitter/receiver circuitry.

efs: Energy consumption for a bit per square meter for free space model.

emp: Energy consumption for a bit per meter to the fourth for multi-path fading model.

For a Cluster Head (CH) of cluster j (for direct transmission):

$$ECHj = Esense + \sum_{i \in CMs\ of\ j} ERx(k) + Eproc(NCMj \cdot k) + ETx\left(kagg, d_{CH_J \to BS}\right) \tag{4}$$

where ERx(k)=k·Eelec;

For multi-hop

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

Eproc for processing/aggregation might be a function of the number of received packets or total data. Residual energy for each node is

$$Eres(t) = Einitial - \sum Econsumed(t) \tag{6}$$

Here, summation of energy consumed of round t, Econsumed *(t)* is depends on the how many cluster is used in the round. The individual node energy loss is depending on whether the node was a cluster head or member during round, and also depend on distance of data transmission. It is important to develop a approach for section of cluster head selection and no of cluster.

3.2. Reinforcement Learning (Q-learning) Framework

Each sensor node acts as an independent RL agent that learns an optimal policy for its role (CH or member).

3.2.1. State Space

The state of an agent (sensor node) is defined by two crucial parameters, discretized into a finite number of states:

I. Residual Energy Level (E_{state}): The current node energy, normalized and categorized into 5 bins. Higher energy nodes are generally more capable of being CHs.
II. Local Node Density (D_{state}): The number of active neighboring nodes within a predefined density Radius. This indicates the compactness of the node's local environment. A higher density might suggest a greater need for a CH in that area to aggregate data.

The combined state S for the Q-table is a unique index derived from these two discrete values:

$S=(E_{state}-1)\cdot\text{numDensityStates}+D_{state}$

3.2.2. Action Space

Each node has two possible actions:

I. Become CH: The node elects itself as a candidate for a Cluster Head.
II. Remain Member: The node chooses to be a regular sensor node, transmitting data to a CH.

3.2.3. Reward Function

The reward function guides the learning process by providing feedback on the desirability of actions. Our reward function is designed to promote energy efficiency and network longevity:

- Positive Reward for Participation: A small positive reward is given for being an active CH or a member, encouraging nodes to contribute to the network.
- Penalty for Node Death: A large negative reward (penalty) is imposed if a node's energy falls below Emin, strongly discouraging actions that lead to premature death.
- Reward for Energy Conservation: A positive reward proportional to the energy conserved (or inversely proportional to energy consumed) in the current round. This directly incentivizes actions that lead to less energy expenditure.

$R_{energy}=(E_{current}-E_{old})\cdot\text{EnergyWeight}$

where $E_{current}$ is energy after consumption, E_{old} is energy before consumption, and EnergyWeight is a scaling factor.

The Q-value update rule follows the standard Q-learning equation:

$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma\ max_A Q(S_t, A_t) - Q(S_t, A_t)]$

where α is the learning rate and γ is the discount factor.

3.2.4. Epsilon-Greedy Policy

An ϵ-greedy policy is used for action selection.

3.3. Cluster Head Selection Mechanism

The process of cluster head selection can be divided into steps that occur in rounds, each round repeating the whole cycle [10,11,12].

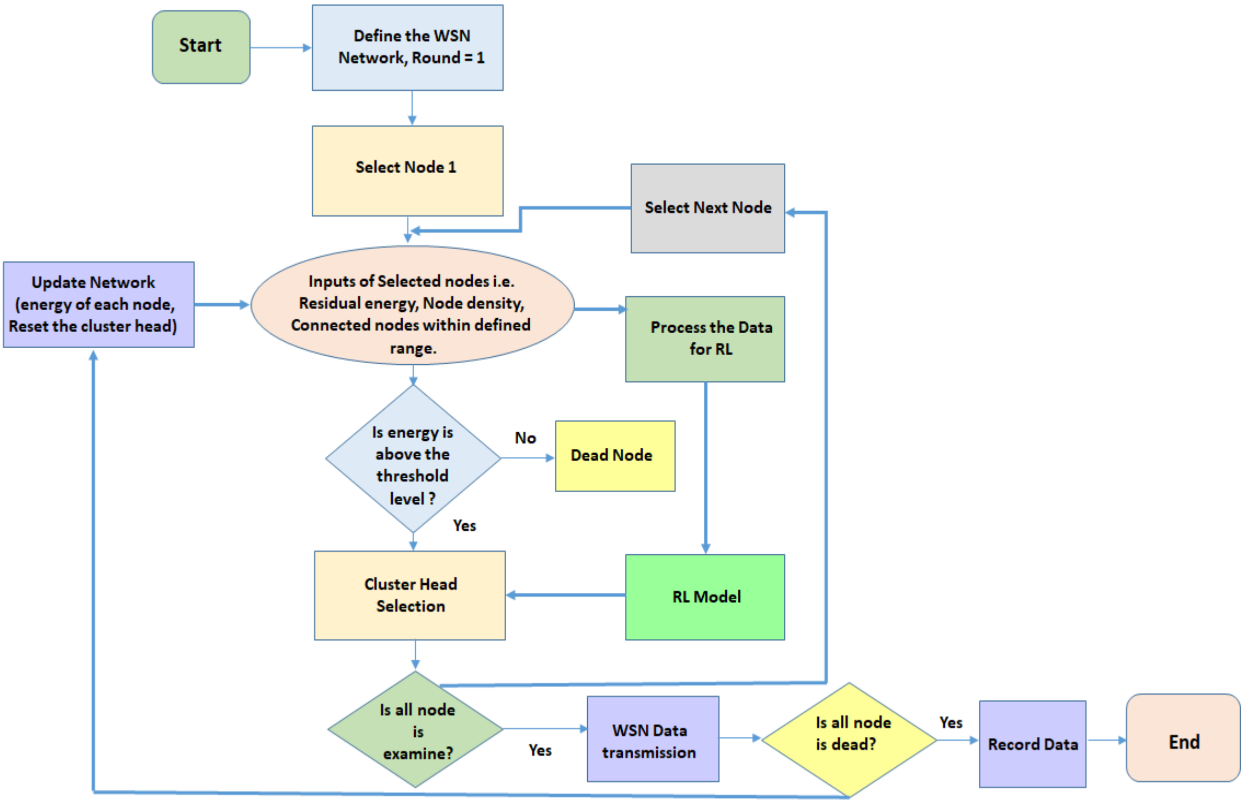
Start
Define the WSN Network, Round = 1
Select Node 1
Select Next Node
Update Network (energy of each node, Reset the cluster head)
Inputs of Selected nodes i.e. Residual energy, Node density, Connected nodes within defined range.
Process the Data for RL
Is energy is above the threshold level ?
No
Dead Node
Yes
Cluster Head Selection
RL Model
Is all node is examine?
Yes
WSN Data transmission
Is all node is dead?
Yes
Record Data
End

Figure 2. Flowchart of cluster head selection in WSN using RL

Step 1: Deployment & Initialization: Sensor nodes are randomly or manually placed in a region. A Base Station (BS) is designated to collect aggregated data from clusters.

Step 2: Cluster Head (CH) Election: Each node activates and probabilistically or energy-based criteria determine if it becomes a CH (e.g., residual energy, fixed probability). The nodes with higher energy are more likely to be selected to balance load.

Step 3: CH Announcement & Node Association: Selected CHs broadcast their status. Nearby nodes listen, evaluate signal strength, and join the closest or strongest CH.

Step 4: Cluster Formation: Nodes respond to their chosen CH, forming clusters with one CH and multiple member nodes.

Step 5: TDMA Scheduling: The CH assigns time slots to members, preventing collisions and reducing energy waste during data transmission.

Step 6: Sensing & Data Transmission: Nodes sense environmental parameters and send data to the CH during their scheduled slots.

Step 7: Data Aggregation: The CH aggregates, filters, and compresses data to reduce size and redundancy before transmission.

Step 8: Sending Data to BS: The CH transmits the aggregated data to the BS, which consumes more energy due to longer distance.

Step 9: Repeat Cycle: Once data is sent, the round ends, and a new CH selection and data collection cycle begins, rotating roles to balance energy use.

3.4. Energy Consumption Simulation

In each round, after actual CHs and members are determined:

- CH Energy: Each selected CH consumes energy for:
    - Receiving data from its assigned cluster members (members connect to their closest active CH).
    - Aggregating the received data.
    - Transmitting the aggregated data to the BS, using the distance-based energy model.
- Member Energy: Each member node consumes energy for:
    - Transmitting its data to its closest active CH. If no CHs are active, members transmit directly to the BS (incurring high energy cost).

Node energies are updated, and if any node's energy falls below 0.01 J, it is marked as dead.

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
- Discount Factor (γ): 0.9
- Initial Epsilon (ϵ): 0.9
- Epsilon Decay: 0.995
- Minimum Epsilon: 0.1
- Simulation Rounds: 5000
- Desired CHs: Max 10% of alive nodes (min 1 CH)

A simulation with the above parameters has been performed with RL-selected CHs for each round separately. The no. of CHs, total alive nodes, total dead nodes, remaining energy of each node and the total energy of the network was recorded. The spatial distribution of the nodes with the base station during simulation is shown in Figure 3.

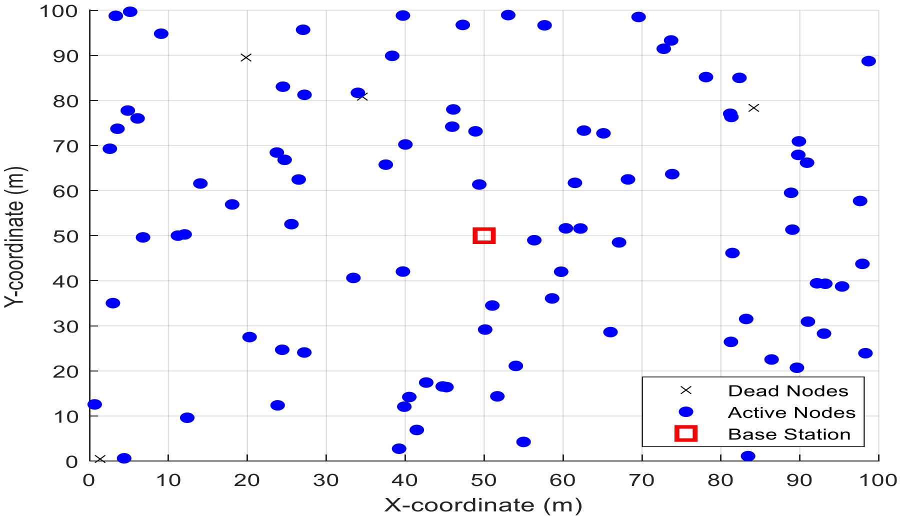

Figure 3. Distribution of nodes (active and dead nodes) along with the base station during simulation.

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

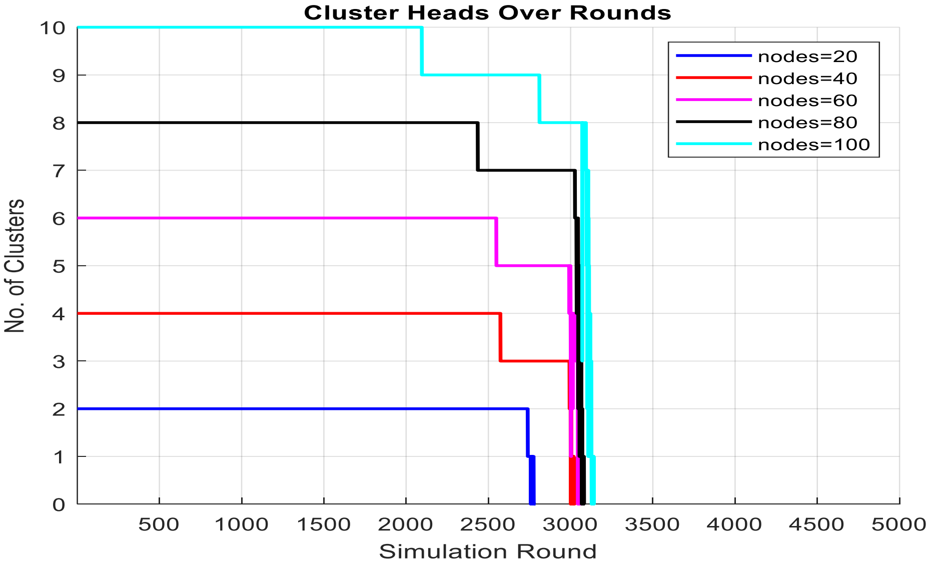

Figure 4. Typical graph for no. of Clusters over Rounds for different no. of Nodes

This graph displays number of clusters over the simulation rounds. The number of clusters for each network size remains consistent throughout the majority of the simulation period. For instance, the 20-node network maintains 2 clusters, while the 100-node network maintains 10 clusters. Similar to the other metrics, a dramatic reduction in the number of clusters occurs for all network sizes around the 3000-round mark.

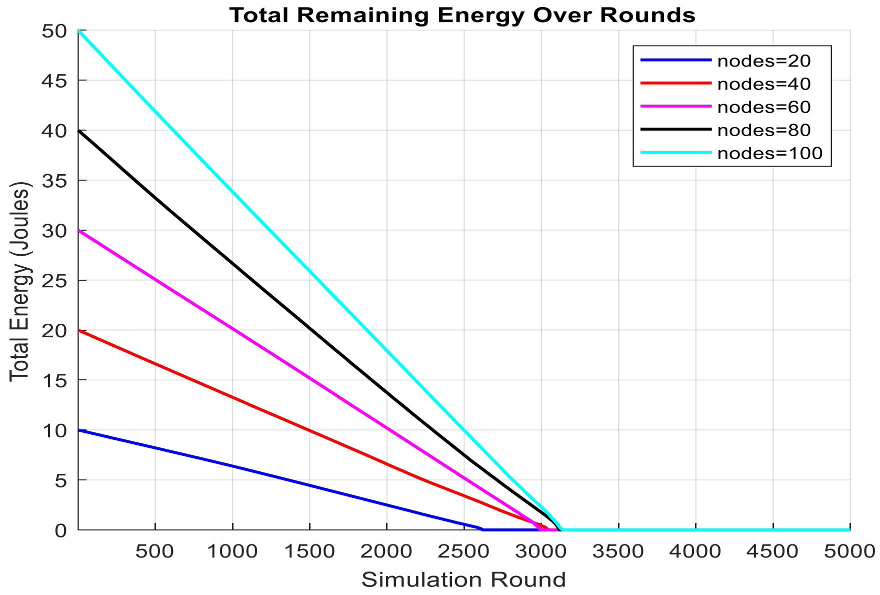

Figure 5. Total Remaining Energy over Rounds

This graph displays the decline in total energy consumption observed in networks comprising 20, 40, 60, 80, and 100 nodes. Each network configuration begins with a different initial total energy, with larger networks possessing more energy. For example, the 100-node network starts with approximately 50 Joules of energy, while the 20-node network begins with around 10 Joules. In all cases, the total remaining energy gradually decreases over the simulation rounds, with a sharp drop-off occurring around the 3000-round mark.

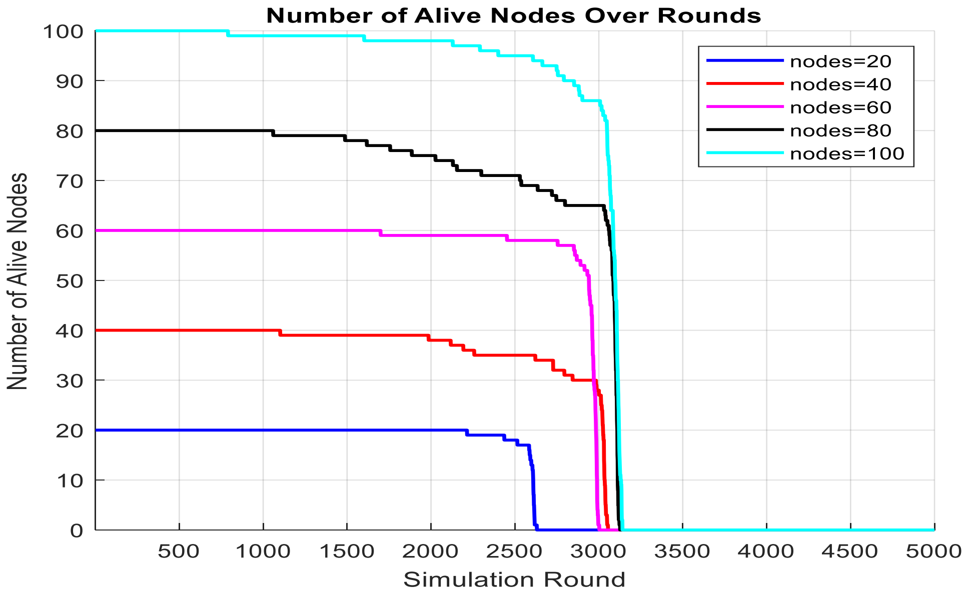

Figure 6. Number of Alive Nodes Over Rounds

This graph shows the decline in total energy consumption observed in networks comprising 20, 40, 60, 80, and 100 nodes. Each network configuration begins with a different initial total energy, with larger networks possessing more energy. For example, the 100-node network starts with approximately 50 Joules of energy, while the 20-node network begins with around 10 Joules. In all cases, the total remaining energy gradually decreases over the simulation rounds, with a sharp drop-off occurring around the 3000-round mark.

The simulation results show that networks with more nodes have higher initial total energy and sustain a larger number of alive nodes and clusters for a longer duration.

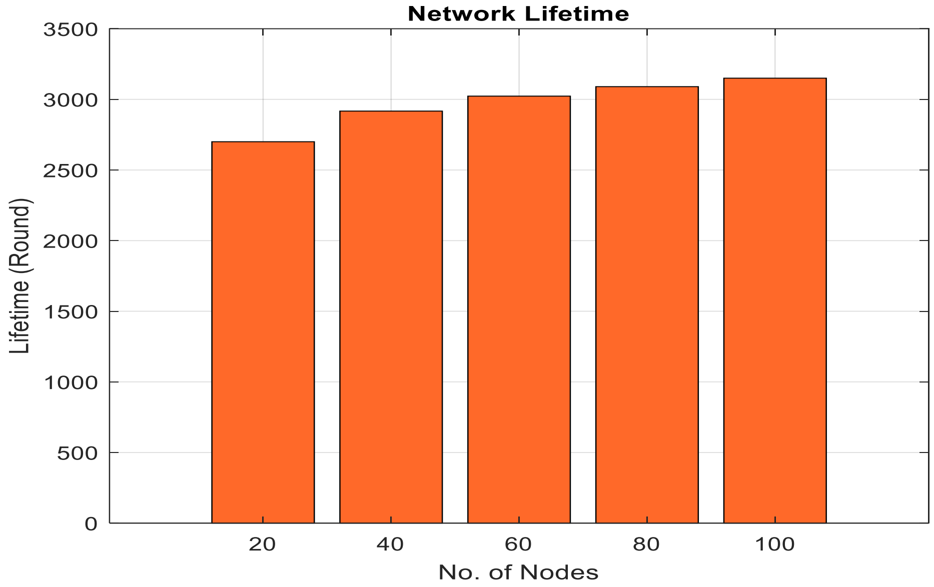

Figure 7. Network Lifetime with No. of Nodes

For all network sizes, there is a consistent decline in total energy over time, accompanied by a sharp drop in both the number of active nodes and the number of clusters, which occurs around the 3000-round mark. The total remaining energy for all node configurations is depleted around the 3000th simulation round.

5 CONCLUSION

This simulation demonstrates a Reinforcement Learning (RL)-based approach for optimizing energy conservation by choosing the most judicious Cluster Head (CH) within WSNs. In the present work, the RL state space is defined with respect to residual energy reserves and localized node density metrics. A reward function is used that incentivizes energy thrift and penalizes premature node attrition; individual sensor nodes acquire learned strategies for assuming their roles. Moreover, the CH selection paradigm strategically exploits these acquired policies, synergized with geospatial proximity to the Base Station, to attenuate cumulative energy expenditure.

There are few limitations of this simulation. Anticipated simulation outcomes are poised to substantiate the efficacy of this adaptive schema in extending network longevity and amplifying energy efficiency. Hence, for further enhancement, future studies could examine more intricate network dynamics, incorporate multi-hop routing paradigms, and investigate the deployment of Deep Reinforcement Learning architectures to accommodate larger-scale and heterogeneous WSN deployments.

ABBREVIATIONS

Symbols	Definition
N	Number of homogeneous sensor nodes
Xmax	Maximum dimension in the X direction
Ymax	Maximum dimension in the Y direction
Nc	Cluster Size (number of nodes within each cluster)
ETx	Energy consumption for transmission (general or for CMs)
ETx1	Energy required for CH to BS transmission
ERx(k)	Energy absorbed by the CH upon reception of k bits
d	Euclidean distance

ECMi	Energy consumed by Cluster Member i in cluster j per round
Esense	Energy consumed for sensing
k	Number of bits
d0	Crossover distance (free space model)
Eelec	Energy consumed per bit by the transmitter/receiver circuitry
efs	Energy consumption for a bit per square meter (free space model)
emp	Energy consumption for a bit per meter to the fourth (multi-path) fading model)
ECHj	Energy consumed by Cluster Head j per round
Eproc	Energy for processing/aggregation
NCMj	Number of Cluster Members in cluster j
kagg	Aggregated data (number of bits)
ETx_path	Energy for transmission over a multi-hop path to the BS
Eres(t)	Residual energy for each node at time t
Einitial	Initial energy
Econsumed(t)	Energy consumed in round t
S	Combined state for the Q-table
E_{state}	Residual Energy Level (normalized and categorized)
D_{state}	Local Node Density (number of active neighbors)
numDensityStates	Number of discrete density states
Emin	Minimum energy level
R_{energy}	Reward for energy conservation
$E_{current}$	Energy after consumption in the round
E_{old}	Energy before consumption in the current round
EnergyWeight	Scaling factor for Renergy
$Q(S_t,A_t)$	Q-value for state St and action At
α	Learning rate
R_{t+1}	Reward received at time t+1
γ	Discount factor
ϵ	Epsilon (probability of exploration)

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest related to this research work.

AUTHOR CONTRIBUTION

Ruchi Sharma: Proposed the research problem, contributed to the conceptualization, design, implementation, experimentation, data collection, analysis, and writing the manuscript.
Sachin patel: Provided supervision, expert guidance, critical review, and valuable suggestions.

REFERENCES

1. Y. Pinar, A. Zuhair, A. Hamad, A. Resit, K. Shiva and A. Omar, Wireless Sensor Networks (WSNs), IEEE Long Island Systems, Applications and Technology Conference (LISAT), 2016, Farmingdale, NY, USA, DOI: 10.1109/LISAT.2016.7494144.
2. S. Zhang and H. Zhang, A review of wireless sensor networks and its applications, IEEE International Conference on Automation and Logistics, 2012, Zhengzhou,China, DOI: 10.1109/ICAL.2012.6308240.
3. Sumana Naskar, Wireless Sensor Networks Challenges and Solutions, IntechOpen,2023, DOI: 10.5772/intechopen.109238.

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, 2000, Maui, HI, USA, DOI: 10.1109/HICSS.2000.926982.
5. S. Lindsey and C. S. Raghavendra, PEGASIS: Power-efficient gathering in sensor information systems, Proceedings, IEEE Aerospace Conference, Big Sky, 2002, MT, USA, DOI: 10.1109/AERO.2002.1035242.
6. O. Younis and S. Fahmy, HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks, IEEE Transactions on Mobile Computing, 2004, DOI: 10.1109/TMC.2004.41.
7. A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, Internet of Things: A survey on enabling technologies, protocols, and applications, IEEE Communications Surveys & Tutorials,2015, DOI: 2347-2376, 2015.
8. Arya, Anju, Reinforcement Learning based Routing Protocols in WSNs: A Survey, International Journal for Research in Applied Science and Engineering Technology,2018, DOI: 3523-3529. 10.22214/ijraset.2018.4584.
9. Tripti Sharma, Archana Balyan, Rajit Nair, Paras Jain, Shivam Arora, Fardin Ahmadi, ReLeC: A Reinforcement Learning-Based Clustering-Enhanced Protocol for Efficient Energy Optimization in Wireless Sensor Networks, Wireless Communication and Mobile Computing,2022, DOI: https://doi.org/10.1155/2022/3337831.
10. A. F. E. Abadi, S. A. Asghari, M. B. Marvasti, G. Abaei, M. Nabavi and Y. Savaria, RLBEEP: Reinforcement-Learning-Based Energy Efficient Control and Routing Protocol for Wireless Sensor Networks, IEEE Access, 2022, vol. 10, pp. 44123-44135, DOI: 10.1109/ACCESS.2022.3167058.
11. S. Mody, S. Mirkar, R. Ghag and P. Kotecha, Cluster Head Selection Algorithm For Wireless Sensor Networks Using Machine Learning, International Conference on Computational Performance Evaluation (ComPE), 2021 Shillong, India, DOI: 10.1109/ComPE53109.2021.975226.
12. Zhang Zhaohui, Zhou Jiaqi, and L. Jing, Q-learning-based semi-fixed clustering routing algorithm in WSNs, Ad Hoc Networks,2025,pp. 103837–103837, DOI: https://doi.org/10.1016/j.adhoc.2025.103837.