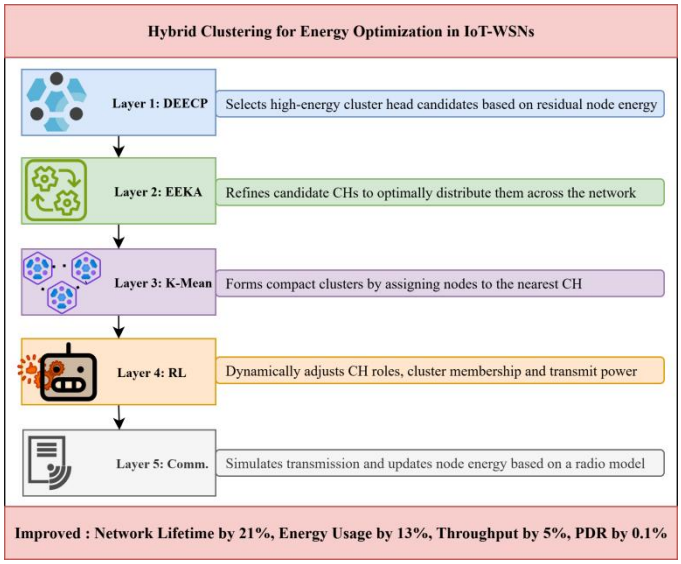
Reinforcement Learning Driven Hybrid Clustering for Energy Optimization in Agri-IoT WSNs

Shubham Kumar
Department of Electronics and Communication
National Institute of Technology Patna, India
shubhamk.pg24.ec@nitp.ac.in

Bharat Gupta
Department of Electronics and Communication
National Institute of Technology Patna, India
bharat@nitp.ac.in

Rakesh Ranjan
Department of Electronics and Communication
National Institute of Technology Patna, India
rr@nitp.ac.in

***Abstract*—In the realm of the Internet of Things (IoT), Wireless Sensor Networks (WSNs) serve as a foundational technology, enabling diverse applications such as urban infrastructure management, industrial automation and environmental monitoring. Despite their widespread adoption, achieving energy efficiency and adaptive clustering remains a major challenge in prolonging the operational lifetime of WSNs. Traditional clustering algorithms such as Low-Energy Adaptive Clustering Hierarchy (LEACH) and Distributed Energy-Efficient Clustering (DEEC) often suffer from uneven energy consumption and static decision-making, limiting their scalability under dynamic network conditions. To address these limitations, this paper proposes a Reinforcement Learning Driven Hybrid Clustering (RLHC) framework that integrates DEEC, the Energy-Efficient Knapsack Algorithm (EEKA) and K-Means with Q-learning based adaptive optimization. In the proposed method, DEEC identifies high energy cluster head (CH) candidates, EEKA ensures energy balanced uniform spatial distribution of CHs, K-Means forms compact clusters to minimize intra-cluster distances and the Q-learning agent dynamically learns optimal adjustment strategies by observing network states defined by residual energy, cluster load and packet delivery ratio (PDR) then executes actions such as CH switching, member reassignment and transmission power tuning. Through continuous interaction with the environment, the agent converges toward energy-optimal configurations. Simulation results demonstrate that the proposed RLHC method significantly enhances network lifetime, PDR and energy balance compared to optimized algorithms built on top of LEACH and DEEC. The improvements include a 21% increase in network lifetime, 13% reduction in energy consumption, 5% higher throughput and 0.1% improvement in PDR. This hybrid intelligence approach provides a scalable and adaptive solution for next-generation Agri-IoT based WSN applications.**



***Keywords*—Reinforcement Learning (RL), Q-learning, Wireless Sensor Networks (WSNs), Distributed energy-efficient clustering (DEEC), Energy Efficient Knapsack Algorithm (EEKA), Energy Optimization, Internet of Things (IoT), K-Means Clustering.**

I. Introduction

Wireless Sensor Networks (WSNs) represent one of the foundational technologies driving the rapid evolution of the Internet of Things (IoT) paradigm. These networks consist of spatially distributed sensor nodes that cooperatively monitor and record environmental conditions and transmit the collected information to a central Base Station (BS) or sink node for further processing and analysis. The ability of WSNs to provide real-time, context-aware data makes them indispensable in modern intelligent systems. They find extensive applications in smart city governance, industrial automation, military surveillance etc [1], [2]. Through efficient data acquisition and communication, WSNs serve as the backbone of IoT ecosystems, enabling data-driven decision-making, predictive analytics and autonomous operations. Despite their vast potential, WSNs face critical constraints in terms of energy efficiency, communication reliability and network longevity. The sensor nodes are typically powered by small, non-rechargeable batteries and deployed in remote or harsh environments, where human intervention for maintenance or battery replacement is highly impractical. Consequently, energy depletion of individual nodes leads to node death, which in turn causes network partitioning, packet loss and degradation in overall system performance. This problem is particularly exist in applications requiring continuous and long-term monitoring such as health diagnostics, disaster prediction and precision agriculture, where uninterrupted operation and consistent data flow are imperative. To mitigate such challenges, researchers have proposed numerous energy aware routing and clustering protocols aimed at optimizing energy utilization and balancing power consumption among nodes. Among these, clustering based routing has emerged as one

of the most effective strategies for achieving energy efficiency. In this approach, sensor nodes are grouped into clusters, each governed by a Cluster Head (CH) that aggregates data from member nodes and transmits it to the BS. Representative clustering protocols include the Low-Energy Adaptive Clustering Hierarchy (LEACH), Stable Election Protocol (SEP) and Distributed Energy-Efficient Clustering (DEEC) [3]–[5]. These protocols primarily aim to reduce communication overhead and distribute the energy load more evenly across the network. DEEC, in particular, selects CHs based on residual energy and average network energy, thereby extending the network lifetime more effectively than random or static CH selection methods.

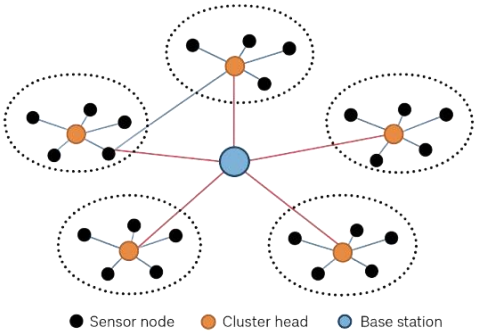


Figure 1: Wireless Sensor Network

Figure 1, illustrates a typical communication model of a WSN, where sensor nodes are randomly deployed across the sensing region and relay their sensed data to the base station through CH.

However, traditional clustering protocols still exhibit limitations in adaptability, scalability and dynamic optimization. Most existing schemes assume homogeneous or quasi-static environments, where node energy dissipation, communication distance and network topology remain relatively stable. In real-world IoT deployments, energy dynamics, node heterogeneity and communication interference can vary significantly over time. As a result, static or semi-static CH selection often leads to sub-optimal cluster configurations, premature energy depletion of certain nodes and overall performance degradation. Additionally, existing methods rarely incorporate adaptive learning or feedback mechanisms that can respond to dynamic environmental conditions or operational uncertainties. Thus the design of an intelligent and adaptive clustering mechanism becomes crucial for achieving long-term sustainability.

To overcome the aforementioned limitations, this paper introduces a Reinforcement Learning-Driven Hybrid Clustering (RLHC) approach that fuses intelligent optimization techniques across multiple layers of the clustering process. The proposed RLHC framework integrates DEEC, EEKA, K-Means clustering and RL to form an adaptive and energy-aware communication architecture. DEEC serves as the foundational layer, prioritizing nodes with higher residual energy for CH selection to ensure balanced energy consumption. EEKA filters and ranks candidate CH based on centrality and spatial distribution, determining the optimal number of CHs and selecting the best CHs to improve spatial balance to reduce inter-cluster communication overhead.. The K-Means algorithm refines the cluster formation by minimizing intra-cluster distances and balancing the spatial distribution of nodes, thereby further reducing communication cost. The RL layer introduces a self-learning capability that dynamically adjusts CH selection, transmission parameters and cluster reconfiguration based on real-time feedback such as residual energy, node density, cluster load and communication delay. The final layer models the radio energy consumption during data transmission and reception between nodes, CHs and the BS using the first-order radio model. This includes the free-space (d^2) and multipath (d^4) channel models for short and long distances, respectively.

By integrating these five layers, the RLHC framework achieves adaptive energy balancing, enhanced throughput, improved PDR and prolonged network lifetime. This hybrid intelligent model demonstrates the potential to bridge the gap between traditional deterministic routing methods and modern ML-driven adaptive optimization techniques. It provides a robust, scalable and self-evolving solution for the future of IoT-enabled WSNs, paving the way for more reliable and sustainable sensor network deployments in real-world environments.

Our contributions can be summarized as follows:

A five-layer RLHC model integrating DEEC (layer 1), EEKA (layer 2), K-Means (layer 3), RL (layer 4) and Communication layer (layer 5) for intelligent energy management in WSNs.

Energy-aware optimal clustering and RL based adaptive cluster refinement mechanism that dynamically optimizes cluster configurations in response to real-time variations in node energy and network conditions.

Performance enhancement as compared to the optimized algorithms built on top of traditional clustering protocols in terms of network lifetime, throughput, energy consumption and packet delivery ratio.

The remainder of this paper is organized as follows:

Section II presents a comprehensive literature review of existing clustering protocols and reinforcement learning approaches in WSNs. Section III outlines the theoretical and technical background relevant to energy-efficient communication and clustering mechanisms. Section IV describes the methodology of proposed Reinforcement Learning-Driven Hybrid Clustering (RLHC) framework in detail which highlighting its five-layer architecture. Section V discusses the simulation environment, performance evaluation metrics and comparative analysis of the proposed RLHC against existing protocols. Finally, Section VI concludes the paper and outlines potential directions for future research and real-world implementation.

II. Related Works

WSNs have become an integral component of modern IoT systems, enabling applications ranging from environmental monitoring and industrial automation to smart cities. A critical challenge in WSNs is energy efficiency, as sensor nodes typically operate on limited battery power and network longevity directly depends on efficient energy utilization. Over the years, researchers have proposed a variety of clustering protocols, optimization algorithms and adaptive learning strategies to address this challenge. This section reviews existing work in this field.

A. Classical Clustering Protocols

Clustering is widely recognized as a foundational approach for energy-efficient WSNs. The LEACH protocol is among the most seminal contributions, introducing randomized rotation of CHs to evenly distribute energy consumption among nodes. It reduces the number of direct transmissions to the base station and organizes nodes into energy-efficient clusters. However, it has several limitations, including uneven CH distribution, lack of consideration for residual energy and poor performance in heterogeneous or large-scale networks. These drawbacks motivated the development of variants that incorporate energy-awareness and residual energy into CH selection. Protocols such as LEACH-RLC and ReLeC enhance LEACH by integrating RL for adaptive CH selection. These methods learn network states and make dynamic decisions about CH election to improve energy efficiency and prolong network lifetime. Despite promising results, RL-based variants can impose computational and communication overhead that may be unsuitable for resource-constrained IoT nodes. Additionally, RL models like ReLeC may face convergence issues in large-scale or highly dynamic WSNs, limiting their practical applicability. While these RL-integrated clustering protocols have demonstrated promising improvements in terms of energy efficiency, stability period, and network throughput, they are not without limitations. The learning process in RL models introduces additional computational and communication overhead, which may not be suitable for resource-constrained IoT nodes with limited processing capabilities. Furthermore, the convergence of RL algorithms such as Q-learning can be slow in large-scale or highly dynamic WSN environments, where the state–action space becomes vast and complex. This can lead to suboptimal decision-making or instability in CH selection, thereby limiting the real-world applicability of such approaches.

B. Heterogeneity-Aware Protocols

Recognizing the limitations of classical homogeneous protocols, researchers have introduced heterogeneity-aware approaches such as SEP (Stable Election Protocol) and DEEC. These protocols consider nodes with different initial energy levels and select CHs based on residual energy and network wide energy distribution. This ensures that higher energy nodes are more likely to assume CH roles, balancing the energy load across the network and preventing premature node failures. DEEC, in particular, extends LEACH by incorporating energy heterogeneity and energy based CH selection, improving network stability and prolonging lifetime. Enhancements such as EEKA further optimize CH selection by considering node centrality and spatial distribution, aiming for a uniform CH spread that reduces intra-cluster communication distances and overall energy consumption. These protocols effectively address energy imbalance but often rely on pre-defined thresholds or heuristic rules, limiting their adaptability to dynamic network topologies or sudden changes in energy states.

C. Metaheuristic and Optimization-Based Approaches

To achieve better CH selection and network optimization, metaheuristic algorithms such as Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO) and Genetic Algorithms (GA) have been extensively applied in WSNs. PSO minimizes intra-cluster distances and identifies optimal CH positions by simulating social behavior among particles. ACO leverages pheromone based path selection to optimize cluster formation and routing, while GA evolves a population of CH candidates to optimize energy efficiency and load balancing. These approaches have demonstrated significant improvements in network lifetime and communication efficiency. However, metaheuristic based methods typically operate offline or rely on iterative convergence, which makes them less suitable for networks with dynamic topologies, mobile nodes or rapidly changing energy states.

D. Fuzzy Logic and Multi-Criteria Clustering

Fuzzy logic-based clustering introduces multi-criteria decision-making for CH selection, considering parameters such as residual energy, node density, distance to the cluster center and network traffic. Protocols like MRCH (Modified RCH-LEACH) utilize fuzzy rules to determine CH candidacy, improving stability, packet delivery ratio and energy efficiency. Fuzzy-based approaches provide a flexible framework for handling uncertainties in sensor networks, enabling adaptive cluster formation under varying network conditions. Moreover, these methods can involve computationally intensive calculations, limiting their deployment on low-power sensor nodes.

E. Reinforcement Learning in WSNs

RL has emerged as a powerful tool for dynamic and adaptive WSN management. RL-based approaches model the network as an environment, where nodes or CHs act as agents that learn optimal actions to maximize long-term rewards, such as energy efficiency or network lifetime. Q-learning, SARSA and Deep RL have been explored to optimize CH selection, cluster reorganization and routing decisions. For example, EER-RL improves energy efficiency and prolongs network lifetime by dynamically adjusting CH roles based on learned energy patterns. Similarly, Q-learning LEACH models enhance adaptability to changing network topologies. Despite their effectiveness, RL-based methods often require centralized training, global knowledge or extensive exploration, which may limit scalability in large or highly dynamic WSN deployments.

F. Hybrid Clustering Approaches

Hybrid clustering methods aim to combine the strengths of classical, meta-heuristic, fuzzy and RL approaches. These methods address multiple challenges simultaneously, including CH optimization, energy balancing, spatial uniformity and adaptability. Multi-layered hybrid models often integrate energy-aware CH selection, K-means or fuzzy-based spatial clustering and RL-based adaptive decision-making to maximize network efficiency. Recent work has highlighted the efficacy of such hybrid frameworks. Protocols like EOCGS determine the optimal number of cluster and grid heads to balance energy consumption. These studies show that intelligent hybrid methods can outperform traditional approaches in terms of network lifetime, energy balance and adaptability.

Despite substantial progress in WSN clustering, existing methods continue to face several challenges. RL based and other meta heuristic approaches often introduce significant computational overhead, making them unsuitable for low-power IoT nodes. Many protocols also struggle to maintain efficiency in large-scale or highly heterogeneous networks. Classical methods frequently fail to prevent early node depletion, resulting in network partitioning. Furthermore, most existing solutions rely on static or pre-defined strategies, which are ill-equipped to handle dynamic topologies, node mobility or sudden energy fluctuations.

TABLE 1. Summary of energy efficiency studies in WSN

Sl. No.	Paper & Year	Optimization Technique Used	Limitations (Research gap)
1	Farahzadi et al. "An Improved Cluster Formation Process in Wireless Sensor Networks to Decrease Energy Consumption" (2021)	Region-based clustering with adaptive CH selection	Assumes ideal energy estimation Does not consider node location or distance factors
2	Panchal et al. "EEHCHR: Energy Efficient Hybrid Clustering and Hierarchical Routing for Wireless Sensor Networks" (2021)	Hybrid clustering with hierarchical routing	Increased routing complexity Static clustering radius
3	Al-Kaseem et al. "Optimized Energy-Efficient Path Planning with Multiple Mobile Sinks" (2021)	Stable Election Algorithm (SEA) Residual energy	Scalability issues beyond 100 nodes Simulation-only
4	Prajapati et al. "Performance Analysis of LEACH with Deep Learning in Wireless Sensor Networks" (2022)	CNN-based CH selection LEACH	High computational overhead Limited scalability.
5	Mohapatra et al. "Mobility Induced Multi-Hop LEACH Protocol in Heterogeneous Mobile Network" (2022)	Residual energy and node mobility factor	Assumes uniform mobility Limited scalability
6	Gamal et al. "Enhancing Lifetime of WSNs Using Fuzzy Logic LEACH and PSO" (2022)	Fuzzy rules, residual energy, node centrality, distance to BS	Increased control overhead from fuzzy inference
7	Bhatia et al. "Cluster Based Energy Efficient Routing Protocol using SA-LEACH to Wireless Sensor Networks" (2023)	Simulated Annealing and LEACH	High computational cost Limited adaptability
8	Abose et al. "Improving Wireless Sensor Network Lifespan with Optimized Energy-Conscious Routing" (2024)	Optimized energy-conscious routing	Assumes ideal energy estimation Does not consider node location or distance factor
9	El Khediri et al. "Energy-Efficient Cluster Routing Protocol for Wireless Sensor Networks" (2024)	Cluster based routing	High computational cost Limited adaptability
10	Zhu et al. "Improved Soft-k-Means Clustering Algorithm for Balancing Energy Consumption in Wireless Sensor Networks" (2024)	Soft-k-means clustering with multi cluster heads	High computational cost Limited adaptability
11	Tabatabaei et al. "New Energy Efficient Management Approach for Wireless Sensor Networks" (2025)	Hierarchical clustering model	Assumes ideal energy estimation Does not consider node location or distance factors.

III. TECHNICAL BACKGROUND

A. Wireless Sensor Networks and Energy Constraints

WSNs are composed of spatially distributed sensor nodes that monitor physical or environmental conditions such as temperature, pressure, humidity or vibrations, and transmit the collected data to a central sink. Each sensor node typically has limited energy, computation capability and communication range, making energy-efficient operation critical to prolong network lifetime and ensure reliable data delivery.

The network operates in rounds, each consisting of cluster formation, data aggregation and transmission. Efficient energy management is critical because sensor nodes are battery powered and recharging may be impractical. Key challenges in WSNs include:

a) Limited energy resources.

b) Uneven energy depletion due to repeated CH selection.

c) Scalability for large networks.

d) Adaptability to dynamic conditions such as node failures, mobility and environmental changes.

To overcome these challenges, cluster based routing is commonly employed, complemented by optimization techniques that efficiently manage intra cluster communication and CH selection, thereby enhancing the overall network lifetime.

B. Particle Swarm Optimization (PSO)

To mitigate the above mentioned limitations various algorithms are developed in which POS is very renowned. It is a population-based optimization algorithm inspired by the social behavior of bird flocking or fish schooling. Each individual in the population called a particle which represents a potential solution. Particles "fly" through the search space, adjusting their positions based on their own best experience and the best experience among all particles. Over iterations, particles converge toward the best solution.

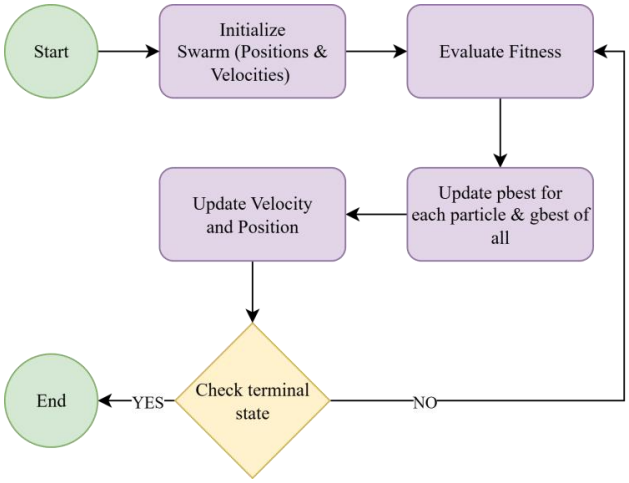


Figure 2: Flow chart of the PSO algorithm

Figure 2 represents the working principle of the Particle Swarm Optimization (PSO) algorithm. The process begins with the initialization of a swarm of particles, each representing a potential solution with random position and velocity. At each iteration, the fitness of all particles is evaluated using the objective function. Each particle then updates its personal best position (pbest) and the global best position (gbest) found by the entire swarm. The velocity and position of each particle are updated according to the Equation 1 and Equation 2 respectively.

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot (pbest_i - x_i(t)) + c_2 \cdot r_2 \cdot (gbest - x_i(t)) \quad (1)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (2)$$

where 'w' denotes the inertia weight, 'c1' and 'c2' are acceleration coefficients and 'r1', 'r2' belongs to [0,1] are random numbers. These Equations collectively balance the inertia (momentum), cognitive (self-learning) and social (swarm cooperation) components of each particle. The process iteratively continues until the swarm converges toward an optimal or near-optimal solution based on the defined fitness function.

C. Ant Colony Optimization (ACO)

Ant Colony Optimization (ACO) is a remarkable metaheuristic algorithm introduced by Marco Dorigo, based on the foraging behavior of real ants. In nature, ants find the shortest path between their colony and a food source by depositing a chemical substance called pheromone on the ground. Other ants sense this pheromone trail and are more likely to follow stronger trails. Over time, shorter paths accumulate more pheromones, leading the colony to converge to the optimal path.

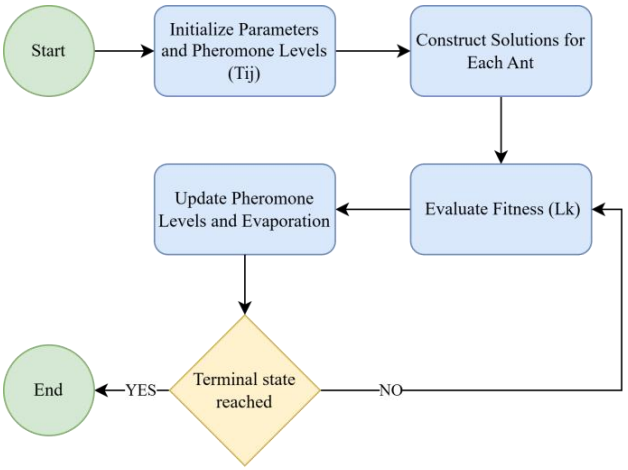


Figure 3: Flow chart of the ACO algorithm

Figure 3 represents the working principle of the Ant Colony Optimization (ACO) algorithm, which simulates artificial ants that iteratively construct solutions using pheromone trails and heuristic information. The process begins with the initialization of parameters and pheromone levels on all paths. During solution construction, each ant builds a solution based on the probability of selecting the next path. After all ants complete their paths, the pheromone update phase reinforces paths used by better solutions while allowing pheromone evaporation to prevent premature convergence. The process continues iteratively until the termination condition such as reaching the maximum number of iterations or convergence is satisfied. The core Equations of ACO are the path selection probability

(Equation 3), pheromone update (Equation 4) and pheromone deposition (Equation 5).

$$P_{ij}^{k}(t) = \begin{cases} \dfrac{[\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\sum_{l \in N_i^k} [\tau_{il}(t)]^{\alpha} \cdot [\eta_{il}]^{\beta}}, & \text{if } j \in N_i^k \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where P_{ij} is probability that ant k moves from node i to node j, T_{ij} is pheromone concentration on edge (i,j) at time t. ηij is equals to 1/dij that is heuristic information (inverse of distance or cost). α, β are control parameters for pheromone and heuristic influence.

$$\tau_{ij}(t+1) = (1-\rho) \cdot \tau_{ij}(t) + \sum_{k=1}^{m} \Delta\tau_{ij}^{k}(t) \quad (4)$$

Equation 4 represents pheromone update rule, where ρ represents pheromone evaporation rate (0<ρ<1), m is total number of ants and ΔT_{ij} amount of pheromone deposited by ant k on edge (i,j).

$$\Delta\tau_{ij}^{k}(t) = \begin{cases} \dfrac{Q}{L_k}, & \text{if ant } k \text{ uses edge } (i,j) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Equation 5 represents pheromone deposition amount, where Q is pheromone constant and L_k is total cost or path length of the tour constructed by ant k.

D. Fuzzy C-Means (FCM)

The FCM algorithm is a well known clustering technique where each data point belongs to a cluster with a degree of membership rather than belonging entirely to just one cluster. This algorithm is fundamentally an optimization based clustering method which seeks to minimize an objective function that quantifies the total weighted distance between data points and cluster centers. Each data point is assigned a membership degree to all clusters, allowing for soft clustering where points can partially belong to multiple clusters. The optimization process aims to find the optimal cluster centers that minimize the objective function by updating the membership degrees for each data point.

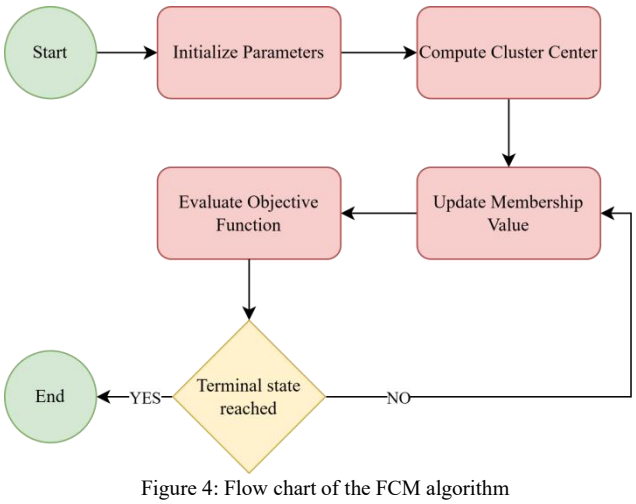


Figure 4: Flow chart of the FCM algorithm

Figure 4 illustrates the working principle of the FCM algorithm. The process begins with the initialization of the membership matrix U, which assigns random membership values to each data point for all clusters. Next, cluster centers are computed based on the current membership values. The algorithm then updates the membership degrees for each data point using the updated cluster centers. Afterward, the objective function is evaluated to measure clustering performance. If the change in membership values or cluster centers between iterations is smaller than a predefined threshold, the process is said to have converged, otherwise, the algorithm repeats the update steps. Once convergence is achieved, the final cluster assignments are generated, representing the optimal fuzzy partition of the dataset.

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \|x_i - c_j\|^2 \quad (6)$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \quad (7)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (8)$$

Equation 6 represents the objective function that FCM seeks to minimize. It measures the total weighted distance between all data points xi and the cluster centers cj. The weight is given by the membership value Uij, m which indicates how strongly a data point belongs to a particular cluster. Equation 7 is used to update the cluster center. Each cluster center is calculated as the weighted mean of all data points, where the weights are the membership degrees raised to the power m. Equation 8 defines how the membership value Uij of each data point to each cluster is updated. The membership is inversely related to the distance between the data point and the cluster center meaning closer points have higher membership values. The ratio term ensures that all membership values for a data point sum to 1 across all clusters, maintaining normalization.

IV. The Methodology

This paper presents a Reinforcement Learning Driven Hybrid Clustering (RLHC) framework to enhance energy efficiency and operational longevity in WSNs. The framework features a five-layer architecture that optimizes CH roles and network management through real-time feedback. The integrated RLHC framework delivers improved throughput, higher packet delivery ratio (PDR), and extended network lifetime by merging deterministic optimization with machine learning driven adaptation, making it robust across diverse IoT enabled WSN scenarios. The RLHC framework comprises five integrated layers designed to enhance energy efficiency and clustering in WSNs.

A. Layer 1: Distributed Energy Efficient Clustering Protocol (DEECP)

DEECP serves as the foundational layer of the proposed RLHC framework. DEECP employs an energy-aware mechanism to identify the most suitable candidates for CH selection by evaluating each node's residual energy relative to the network's average energy. This adaptive CH election strategy ensures that nodes with higher remaining energy possess a greater probability of becoming CHs, thereby balancing energy consumption across the network. Through this dynamic mechanism, DEECP effectively mitigates premature node death and enhances network stability by rotating CH roles periodically among eligible nodes. It not only optimizes the initial CH election but also synergizes with higher level adaptive modules to achieve prolonged network lifetime, balanced load distribution and enhanced overall energy efficiency in heterogeneous IoT-enabled WSN environments.

$$T(i) = \begin{cases} \frac{P_i}{1 - P_i \cdot (r \bmod \left(\frac{1}{P_i}\right))}, & \text{if } i \in G \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

$$P_i = prop \times \left(\frac{E_i}{\bar{E}(r)} \right) \quad (10)$$

$$\bar{E}(r) = \frac{E_{\text{total}} \cdot \left(1 - \frac{r}{R}\right)}{N} \quad (11)$$

Equation 9, represents T(n) which is threshold for node n to become CH, Pi is probability of node i becoming a CH, r is current round number and G is set of nodes that have not been CHs in the last 1/Pi rounds. In Equation 10, the Pi is proportional to the Ei which represents residual energy of node i, E(r) which represent average residual energy of the network at round r, 'prop' is optimal CH probability, N represents total numbers of nodes, R represents total numbers of rounds and Etotal represents total energy of the network.

B. Layer 2: Energy Efficient Knapsack Algorithm (EEKA)

Following the DEECP layer, the EEKA operates as the second optimization layer within the RLHC framework. While DEECP ensures energy aware CH selection based on residual energy and average network energy, EEKA refines this process by optimizing the spatial distribution and number of CHs using a constrained knapsack formulation. In this layer, each potential CH candidate identified by DEECP is evaluated as an item in the knapsack problem, where parameters such as node centrality, residual energy, and communication distance to the base station act as utility factors. It aims to maximize the overall network utility under the constraint of minimizing total energy consumption. By selecting the most spatially balanced and energy-efficient CH set, it also ensures uniform coverage of the sensing field and reduces excessive intra cluster communication distances. This selection prevents the formation of energy hotspots and promotes equitable energy dissipation across all regions of the network. The integration of EEKA into the RLHC framework enhances both structural balance and energy uniformity, addressing the spatial irregularities often observed in conventional DEECP or LEACH based clustering. The resulting CH configuration not only reduces redundant transmissions but also improves communication reliability and network throughput. Furthermore, the optimal CH distribution established by EEKA serves as an informed input to the subsequent K-Means clustering layer, which further fine-tunes cluster boundaries to minimize intra cluster distances and transmission costs.

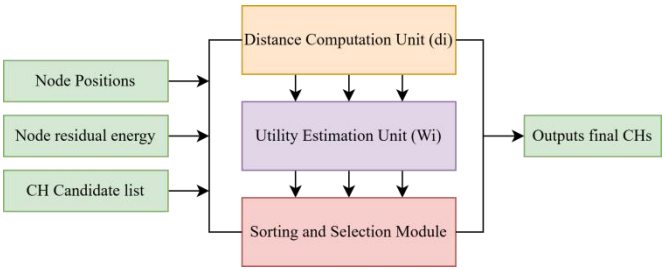


Figure 5: Internal block diagram of the EEKA algorithm

The internal architecture of EEKA operates by receiving candidate cluster heads, along with node energy and positional data. It computes average inter-node distances to estimate node centrality, then evaluates a combined utility function (as defined in Equation 12) to rank the candidates. Finally, the top K nodes are selected as cluster heads, ensuring optimal energy balance and uniform cluster distribution across the network.

$$W_i = E_i + \frac{1}{\bar{d}_i + \epsilon} \quad (12)$$

where the utility function Wi represents the energy centrality score of each node and determines its suitability to become a CH. It integrates two crucial parameters, residual energy and spatial centrality into a single quantitative metric.

C. Layer 3: K-Means Algorithm

The third layer of the RLHC framework employs the K-Means clustering algorithm to refine the cluster formation process based on the CHs selected by the EEKA layer. While DEECP and EEKA collectively determine the most energy efficient and spatially balanced CHs, the K-Means algorithm ensures that the remaining sensor nodes are optimally associated with these CHs to minimize intra cluster communication cost.

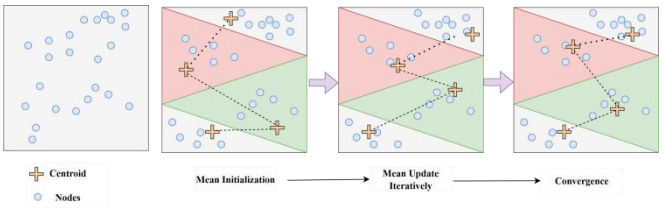


Figure 6: Process flow chart of K-Mean algorithm

The K-Means algorithm partitions a set of N sensor nodes into K clusters by iteratively minimizing the sum of squared Euclidean distances between each node and its assigned cluster head. The process involves three main steps: centroid initialization, mean (centroid) update, and convergence check. The objective function for this clustering process is given in Equation 13.

$$J = \sum_{i=1}^{K} \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (13)$$

where Ci represents the set of nodes in cluster i, μi is the centroid (cluster head) of cluster i, and sq(x−μi) denotes the squared Euclidean distance between node x and its cluster centroid. J is non-convex, meaning K-Means may converge to a local minimum, not necessarily the global one.

D. Layer 4: Reinforcement Learning (RL)

Layer 4 incorporates a RL agent that provides self-learning, adaptive control for the wireless sensor network. It dynamically optimizes CH selection, transmission parameters, and cluster reconfiguration based on real-time network feedback. By employing the Q-Learning algorithm, this layer improves CH selection and transmission decisions to maximize long-term objectives, such as network lifetime and energy efficiency. The RL agent continuously refines its strategies over time, overcoming the limitations of static, rule-based systems. It synergizes with outputs from preceding layers to enhance overall decision-making, ensuring scalable, efficient, and adaptive clustering. While it introduces minimal computational overhead, this layer is essential for enabling intelligent and resilient operation in IoT-based WSN applications.

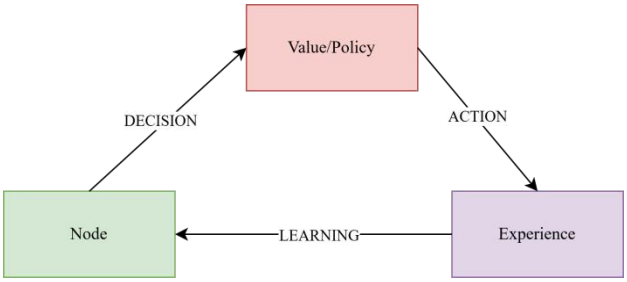


Figure 7. RL based Decision-Making Framework

From the above diagram, it is evident that by incorporating RL, the system can dynamically adapt to changing conditions and make increasingly optimal decisions over time, making it highly suitable for real-time IoT applications.

In Layer 4, a Q-learning agent is deployed whose core function is to observe the current network state, execute actions such as CH reassignment, switching CHs, or adjusting transmission power, and gain experience in the form of rewards. After executing an action, it receives feedback from the environment, including updated node energies, cluster configurations, and network metrics (e.g., PDR, throughput) and computes a reward. This experience is then used to update the Q-table, refining future decision making. By introducing adaptive learning, this layer transforms static cluster management into a self-learning mechanism, enabling the network to optimize long-term objectives such as energy efficiency and network lifetime.

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right] \quad (14)$$

Equation 14 is used to update the Q-table, where α is the learning rate that determines how much new information overrides old knowledge, γ is the discount factor that defines the importance of future rewards, and 'a' represents the action taken by the agent. This update allows the agent to iteratively improve its policy by balancing immediate rewards with long-term benefits, enabling more optimal decision making over time.

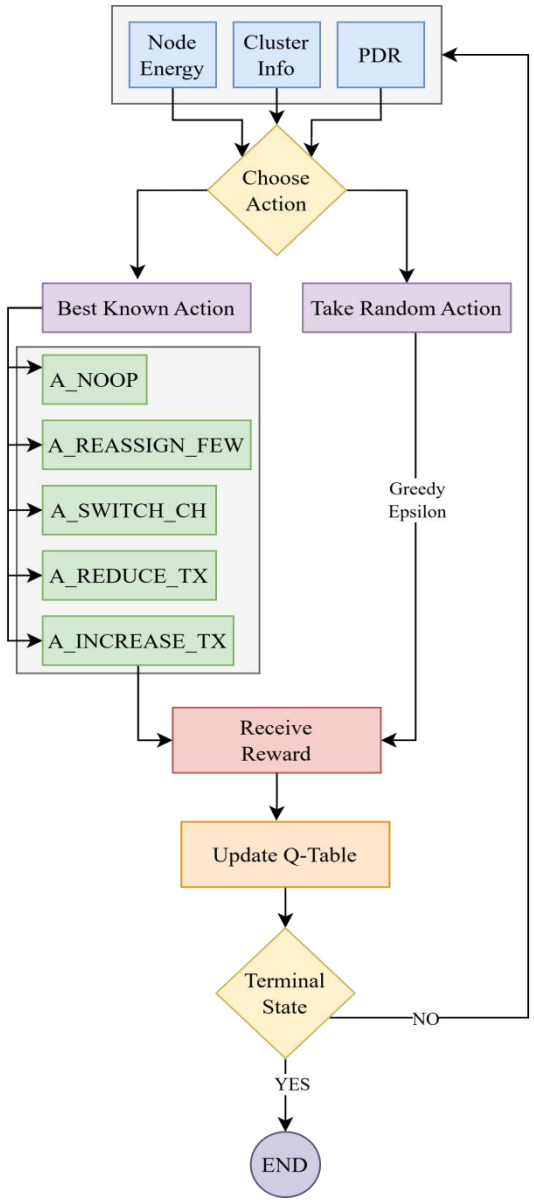


Figure 8. Q-Learning agenrt decision making framework

Figure 8 illustrates how the agent interacts with the environment to make adaptive chnages and fine tune CHs selection, Tx Rx power adjustment and cluster member reassignment decisions.

E. Layer 5: Communication Layer

Layer 5 is responsible for wireless communication and energy modeling during data transmission between nodes, CHs and the BS. This layer simulates the radio energy dissipation model, which estimates the transmission and reception energy costs based on the distance between nodes and the amount of data transmitted. In each communication round, member nodes transmit sensed data to their respective CHs, which then perform data aggregation and forward the aggregated packet to the BS using a single-hop communication mechanism. In the proposed model, wireless communication is simulated using the first-order radio energy model that is Additive White Gaussian Noise (AWGN) channel., which incorporates both free-space (d^2) and multipath (d^4) channel models. For short-range intra-cluster communication ($d<d_0$), the free-space model is used, assuming line-of-sight propagation. For long-range

transmissions ($d \geq d_0$), the multipath fading model is applied to account for non-line-of-sight energy loss. This dual channel approach effectively captures large-scale path loss while maintaining computational simplicity, providing a realistic approximation of WSN communication behavior.

$$E_{Tx}(k,d) = \begin{cases} k \cdot E_{elec} + k \cdot \varepsilon_{fs} \cdot d^2, & \text{if } d < d_0 \\ k \cdot E_{elec} + k \cdot \varepsilon_{mp} \cdot d^4, & \text{if } d \geq d_0 \end{cases} \quad (15)$$

Equation 15, computes the energy needed to transmit a k-bit packet over a distance d, Longer distances ($d \geq d_0$) incur significantly higher energy costs due to multi path fading, where k is size of the data packet in bits, E_elec is energy required per bit for transmission/reception circuitry, ε_fs is amplifier energy for free-space model, ε_mp is amplifier energy for multipath model, d is distance between sender and receiver, d_0 is threshold distance to switch between free-space and multi path.

$$E_{Rx}(k) = k \cdot E_{elec} \quad (16)$$

$$E_{DA}(k) = k \cdot E_{DA} \quad (17)$$

Equation 16, calculates the energy required to receive k bits, receiving costs are independent of distance, as no amplification is needed. Equation 17, E_DA is energy required for data aggregation for k bits before transmission.

$$E_{CH} = \left(\frac{N}{k} - 1\right) kE_{elec} + N \cdot E_{DA} + kE_{elec} + k\varepsilon_{fs} d_{BS}^2 \quad (18)$$

$$E_{nonCH} = k \cdot E_{elec} + k \cdot \varepsilon_{fs} \cdot d_{CH}^2 \quad (19)$$

Equation 18 and 19 calculates the total enegy consume by the CHs and CMs, when considering no multi path fading.

The proposed architecture aims to improve the energy efficiency and communication reliability of Wireless Sensor Networks (WSNs) in Smart Agriculture (Agri-IoT) environments. It employs a hierarchical multi-layer design that combines traditional clustering mechanisms with intelligent reinforcement learning-based optimization. In this framework, a large number of heterogeneous sensor nodes are randomly deployed across the agricultural field to continuously monitor environmental parameters such as soil moisture, humidity, and temperature. The sensed data is processed and transmitted to a centralized BS for decision making and analytics. The architecture is organized into five functional layers. In the first layer, nodes are initialized and deploy DEECP to identifie suitable candidates for CH. The second layer filters out suitable CH candidates by evaluating both residual energy and spatial centrality. The third layer applies the K-Means clustering algorithm to associate non-CH nodes with the nearest CHs, thus minimizing intra-cluster communication distance and balancing energy consumption. The fourth layer integrates a Q-learning agent that intelligently optimizes CH selection based on dynamic network states. The agent observes the environment, selects actions such as reassigning CHs or adjusting transmission power, and updates its Q-table based on the obtained rewards, thereby enabling adaptive and self-learning CH selection over time. Finally, the fifth layer handles data transmission and aggregation, where member nodes send their data to CHs, which then forward aggregated packets to the BS. The radio communication model used in this work incorporates both free-space (d^2) and multipath (d^4) channel propagation characteristics, which are representative of realistic agricultural field conditions where line-of-sight and non-line-of-sight transmissions coexist. This integrated design ensures balanced energy consumption, extended network lifetime, and high data reliability, making it suitable for large-scale and dynamic Agri-IoT deployments.

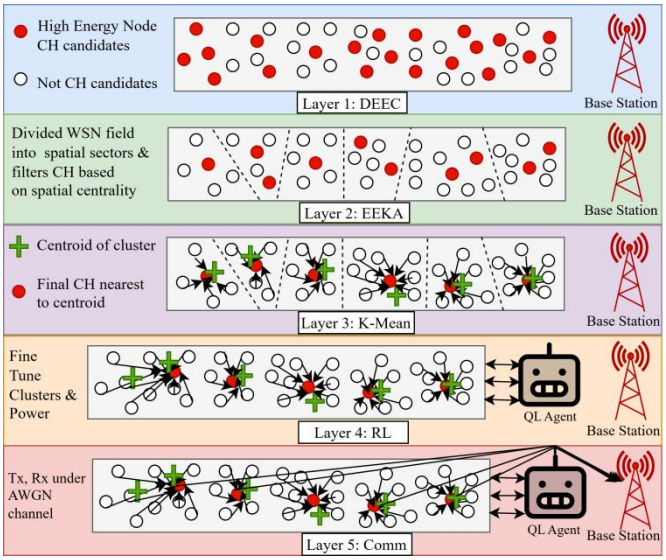


Figure 9. System architecture of the proposed algorithm

Algorithm 1: RHLC - Reinforcement Learning Driven Hybrid Clustering

Initialize:

Network parameters, node positions and energy.
Q-Table Parameters

Set round counter r = 0.

While (any node is alive):

Layer 1 (DEEC): *Estimate CH probability based on residual energy ratio.*

Layer 2 (EEKA): *Divide WSN field into spatial sectors and select energy-efficient CH candidates using node centrality and residual energy.*

Layer 3 (K-Means)*: Form compact clusters by associating nodes with the nearest CH closest to the centroid.*

Layer 4 (Q-Learning): *Observe state (residual energy, distance, alive ratio, pdr, cluster info) Then choose action using ε-greedy policy (CH reassignment / maintain). Then update Q-table using Bellman update equation.*

Layer 5 (Communication): *Transmit data: Member nodes to CH (Free-space, d^2), CH to BS (Multipath, d^4) under AWGN channel. Then update node energies and performance metrics (PDR, throughput etc).*

Increment round counter r.

End While.

Table 2: WSN parameters for Agri-IoT WSNs

Network Parameters	Value
WSN area size	500m x 500m
Number of nodes	200
Initial energy	2 Joules
Packet size	512 bytes
Data aggregation energy	5n J/bit/signal
Transmit energy	50n J/bit
Receive energy	50n J/bit
Free space loss	10n J/bit
Multipath loss	0.0013p J/bit
Simulation Rounds	1000

The simulation will evaluate WSN performance by tracking the number of dead nodes, network lifetime, energy consumption, throughput and improvement in PDR.

V. Results and Discussion

This section presents the performance evaluation results of the optimized algorithms built on top of LEACH and DEEC and the proposed mechanism. The analysis focuses on the network lifetime, energy consumption, throughput and packet delivery ratio. The simulations assume that all nodes are either stationary or exhibit only micro-mobility and energy losses due to dynamic random channel conditions and fading effects are neglected.

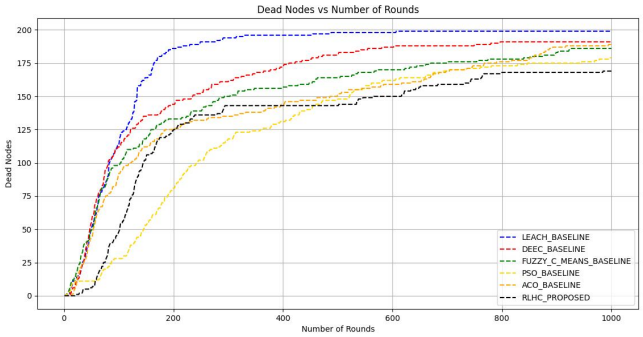


Figure 10. Number of dead nodes during network operation

Figure 10 illustrates the comparative performance of various protocols in terms of network lifetime. The LEACH protocol exhibits the earliest network degradation, with the first node dying around round 60. DEEC offers a slight improvement, extending the first node death to approximately round 100. Fuzzy C-Means and PSO further enhance the network's stability, while ACO demonstrates even better performance. In contrast, the proposed RLHC protocol achieves the longest network lifetime, with the first node death occurring around round 150, representing a 3.8 times improvement over LEACH and approximately 1.2 times over ACO. The slower node death rate observed in RLHC highlights its superior energy balancing and adaptive learning capabilities.

Figure 11 depicts the cumulative throughput trends for all protocols. LEACH achieves the lowest total throughput of approximately 23,000 packets by the end of 1000 rounds due to its premature node failures. DEEC and Fuzzy C-Means reach around 44,000 and 55,000 packets, respectively, benefiting from better cluster head selection strategies. PSO and ACO protocols further enhance throughput, reaching approximately 53,000 and 71,000 packets, respectively. The proposed RLHC protocol achieves the highest cumulative throughput, delivering nearly 75,000 packets to the base station. This 226% increase over LEACH and 5% improvement over ACO demonstrates the effectiveness of reinforcement learning in maintaining optimal network operation and minimizing communication losses.

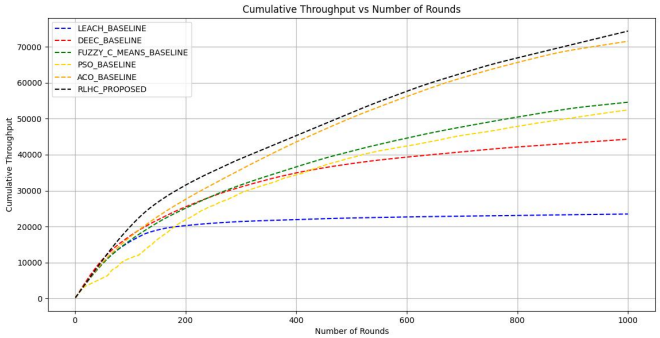


Figure 11. Cumulative throughput during network operation

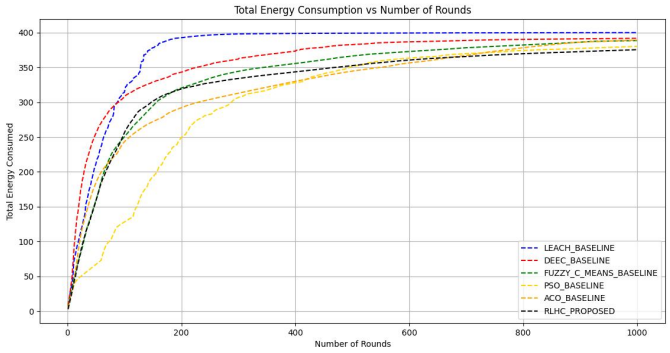


Figure 12. Total energy consumption during network operation

Figure 12 analysis shows how efficiently each protocol utilizes the available energy resources. LEACH and DEEC consume energy rapidly, reaching approximately 400 J total consumption by round 300, leading to early node deaths. Fuzzy C-Means and PSO consume energy more gradually, stabilizing around 370 to 380 J by the end of the simulation. ACO demonstrates further improvement, reaching around 385 J after 900 rounds. The proposed RLHC protocol achieves the lowest overall energy consumption, approximately 360 J after 1000 rounds. This reflects more balanced transmission power control and adaptive energy-aware CH selection.

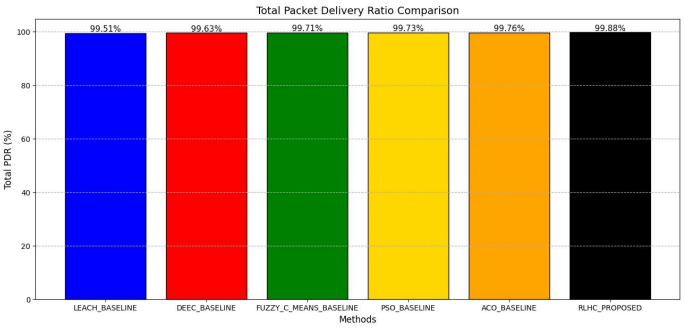


Figure 13. Total PDR during network operation

Figure 13 illustrates the comparative analysis of the Total Packet Delivery Ratio (PDR) achieved by different routing algorithms, namely LEACH, DEEC, Fuzzy C-Means, PSO, ACO, and the proposed RLHC method. The results clearly

indicate that all protocols maintain a high PDR, reflecting their reliable data delivery capability within the network. However, a noticeable improvement can be observed in the proposed RLHC approach, which achieves the highest PDR of 99.88%, followed by ACO with 99.76%, PSO with 99.73%, Fuzzy C-Means with 99.71%, DEEC with 99.63%, and LEACH with the lowest at 99.51%. The enhanced PDR in the RLHC protocol can be attributed to its reinforcement learning-based adaptive CH selection mechanism, which dynamically optimizes communication paths and minimizes packet loss due to energy depletion or transmission failures.

Figure 13 compares the operational lifetime of the network for the same set of algorithms, measured in terms of the total number of rounds completed before the network becomes non-functional. The proposed RLHC approach achieves the longest operational duration of 7128 rounds

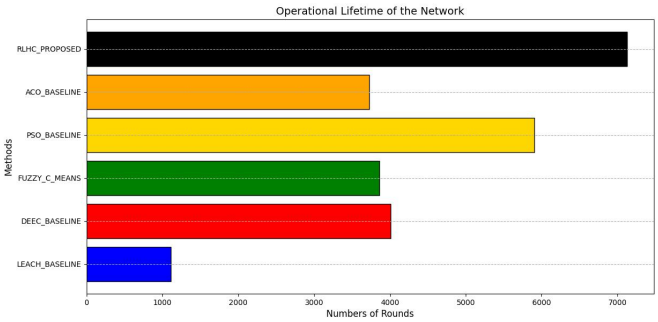


Figure 13. Operational lifetime of the network

outperforming PSO (5904 rounds), DEEC (4008 rounds), Fuzzy C-Means (3862 rounds), ACO (3731 rounds), and LEACH (1110 rounds).

Table 3: Comparison of overall performance

Method	LEACH	DEEC	FCM	PSO	ACO	RLHC (proposed)
Network Lifetime (rounds)	1110	4008	3862	5904	3730	7128
Energy Consumption (J)	400	390	380	380	385	360
Packet Delivery Ratio (%)	99.51	99.63	99.73	99.71	99.76	99.88
Throughput(bits)	23000	44000	55000	53000	71000	75000
First Node Dead (FND)	8	12	4	9	14	21
Half Node Dead (HND)	83	81	103	257	120	145
Last Node Dead (LND)	1110	4008	3862	5904	3730	7128
Energy Model	Homogeneous	Homogeneous	Homogeneous	Hetrogeneous	Hetrogeneous	Hetrogeneous

From the comparative analysis presented in Table 3, it is evident that the proposed RLHC protocol significantly outperforms existing clustering and optimization-based routing techniques such as LEACH, DEEC, FCM, PSO, and ACO across all key performance metrics. RLHC achieves the highest network lifetime (7128 rounds), lowest energy consumption (360 J), and superior packet delivery ratio (99.88%), demonstrating its ability to maintain reliable communication and balanced energy utilization. Furthermore, the delayed occurrence of the first and last node deaths highlights the protocol's enhanced stability and robustness in heterogeneous environments. These results collectively validate the effectiveness of the reinforcement learning driven clustering and adaptive energy management strategies employed in RLHC.

VI. Conclusion

In this study, RLHC protocol was proposed to enhance the energy efficiency and lifetime of heterogeneous wireless sensor networks. By integrating Q-learning with multi-layered optimization mechanisms, RLHC dynamically adapts to network conditions and optimizes cluster head selection and data transmission paths. Simulation results clearly demonstrate that RLHC surpasses traditional protocols such as LEACH, DEEC as well as optimized protocol built on top of traditional algorithm such as FCM, PSO and ACO in terms of network longevity, throughput and energy conservation. The proposed approach effectively balances the energy load among nodes, reduces early node deaths, and sustains overall network connectivity for extended periods.

Future work will focus on extending RLHC to mobile and large-scale sensor networks, incorporating deep reinforcement learning for further adaptability, and validating performance under real-time deployment scenarios to enhance its practical applicability in smart agriculture, smart city and IoT-based systems.

References

[1] S. M. Chowdhury and A. Hossain, ''Different energy saving schemes in wireless sensor networks: A survey,'' Wireless Pers. Commun., vol. 114, no. 3, pp. 2043–2062, Oct. 2020.

[2] F. M. Salman, A. A. Mohammedand A. F. Mutar, "Optimization of LEACH protocol for WSNs in terms of energy efficient and network lifetime," Journal of Cyber Security and Mobility, vol. 12, no. 3, pp. 275–296, May 2023, doi: 10.13052/jcsm2245-1439.1232.

[3] G. M. T. Tamilselvan and K. Gandhimathi, ''Network coding based energy efficient LEACH protocol for WSN,'' J. Appl. Res. Technol., vol. 17, no. 1, pp. 251–267, Jun. 2019

[4] W. Heinzelman, A. Chandrakasanand H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Hawaii Int. Conf. System Sciences*, 2000, pp. 1–10, doi: 10.1109/HICSS.2000.926982.

[5] A. Chehri, R. Saadane, N. Hakemand H. Chaibi, ''Enhancing energy efficiency of wireless sensor network for mining industry applications,'' Proc. Comput. Sci., vol. 176, pp. 261–270, Jan. 2020.

[6] P. Kathiroli and K. Selvadurai, ''Energy efficient cluster head selection using improved sparrow search algorithm in wireless sensor networks,'' J. King Saud Univ.-Comput. Inf. Sci., vol. 2021, pp. 1–12, Sep. 2021.

[7] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," MIT Press, 2nd ed., 2018. [Online]. Available: http://incompleteideas.net/book/the-book-2nd.html

[8] H. Zhang, G. Liuand T. Jiang, "Parameter configuration scheme for optimal energy efficiency in LoRa-based wireless underground sensor networks," IEEE Transactions on Vehicular Technology, vol. 74, no. 6, pp. 9961–9973, Jun. 2025, doi: 10.1109/TVT.2025.3351526.

[9] B. R. Al-Kaseem, Z. K. Taha, S. W. Abdulmajeedand H. S. Al-Raweshidy, "Optimized energy–efficient path planning strategy in WSN with multiple mobile sinks," IEEE Access, vol. 9, pp. 79994–80007, Jun. 2021, doi: 10.1109/ACCESS.2021.3087086.

[10] F. M. Salman, A. A. Mohammedand A. F. Mutar, “Optimization of LEACH protocol for WSNs in terms of energy efficiency and network lifetime,” Journal of Cyber Security and Mobility, vol. 12, no. 3, pp. 275–296, May 2023, doi: 10.13052/jcsm2245-1439.1232.

[11] M. Gamal, N. E. Mekky, H. H. Solimanand N. A. Hikal, “Enhancing the lifetime of wireless sensor networks using fuzzy logic LEACH technique-based particle swarm optimization,” IEEE Access, vol. 10, pp. 36935–36948, Mar. 2022, doi: 10.1109/ACCESS.2022.3163254.

[12] A. Panchal and R. K. Singh, “EOCGS: Energy efficient optimumnumber of cluster head and grid head selection in wireless sensor networks,” Telecommun. Syst., vol. 78, no. 1, pp. 1–13, Apr. 2021, doi:10.1007/s11235-021-00806-w.

[13] R. K. Singh, S. Verma, A. Panchal, and S. Dubey, “Modified RCH LEACH (MRCH) for wireless sensor networks (WSN),” in Proc. 9th Int. Congr. Inf. Commun. Technol. (ICICT). Singapore: Springer, Jan. 2024, pp. 331–340, doi: 10.1007/978-981-97-3559 4_26.

[14] K. Yilmaz, R. Kara, and F. Katircioglu, “Energy-efficient hybrid adaptive clustering for dynamic MANETs,” IEEE Access, vol. 13, pp. 51319–51331, 2025, doi: 10.1109/ACCESS.2025.3552232.

[15] S. Kaviarasan and R. Srinivasan, “A novel spider monkey optimized fuzzy C-means algorithm (SMOFCM) for energy-based cluster-head selection in WSNs,” Int. J. Electr. Electron. Res., vol. 11, no. 1, pp. 169–175, Mar. 2023, doi: 10.37391/ijeer.110124.

[16] S. K. Chaurasiya, S. Mondal, A. Biswas, A. Nayyar, M. A. Shah, and R. Banerjee, “An energy-efficient hybrid clustering technique (EEHCT) for IoT-based multilevel heterogeneous wireless sensor networks,” IEEE Access, vol. 11, pp. 25941–25958, 2023, doi: 10.1109/ACCESS.2023.3254594.