DOI: 10.21917/ijct.2024.0488

ADAPTIVE REINFORCEMENT LEARNING-BASED DATA AGGREGATION AND ROUTING OPTIMIZATION (ARL-DARO) FOR ENHANCING PERFORMANCE IN WIRELESS SENSOR NETWORKS

V. Shobana and Jasmine Samraj

Department of Computer Science, Quaid-E-Millath Government College for Women, India

Abstract

Wireless Sensor Networks (WSNs) are challenged by the need for optimized Energy Consumption (EC), efficient Data Aggregation (DA), and reliable routing due to their dynamic topologies and limited resources. Existing solutions like TEAMR and DDQNDA address these concerns but face significant drawbacks—TEAMR lacks adaptability to rapidly changing topologies, while DDQNDA suffers from high computational overhead and delayed convergence, hindering its effectiveness in real-time scenarios. To overcome these limitations, this paper introduces the Adaptive Reinforcement Learning (RL)-Based DA and Routing Optimization (ARL-DARO) algorithm. The proposed methodology follows a systematic approach, beginning with cluster formation and Cluster Head (CH) selection (CHS) using the Grey Wolf Optimizer (GWO), which ensures Energy-Efficient (EE) clustering and optimal CH selection. In the next step, trust factors such as Node Connectivity (NC), Residual Trust (RT), and Cooperation Rate (CR) are integrated into Quality of Service (QoS) metrics as part of the Fitness Function(FF) to enhance route reliability and security. Finally, the ARL-DARO algorithm is employed to dynamically optimize both data aggregation and routing. It leverages Q-learning to select optimal routes based on energy efficiency, security, and link reliability, further reducing data redundancy and improving adaptability to real-time network changes. Performance is assessed using parameters such EC, packet delivery ratio (PDR), end-to-end latency (E2E delay), throughput, and network lifetime (NL) across networks with 100, 200, 300, 400, and 500 nodes. Results show that ARL-DARO significantly reduces energy consumption by up to 45%, increases throughput by 30%, and extends network lifetime, proving its effectiveness over existing methods.

Keywords:

Wireless Sensor Networks (WSNs), Reinforcement Learning, Routing Optimization, Data Aggregation, Energy Efficiency, Trust-Aware Routing, Adaptive Algorithms

1. INTRODUCTION

Wireless Sensor Networks (WSNs) have emerged as critical technologies for various applications such as environmental monitoring, biomedical health tracking, and target detection [1]. These networks consist of distributed Sensor Nodes (SN) that collaborate to monitor environmental conditions and transmit data to a central sink node. A major challenge in WSNs is extending the network lifetime (NL) while maintaining energy efficiency and secure communication, particularly since SNs are often battery-powered and deployed in hostile environments. Cluster-Based Routing (CBR) has been widely explored as an energy-efficient strategy for extending NL [2], where Cluster Heads (CH) manage communication within clusters and relay aggregated data to the sink. Traditional CBR algorithms, such as Low-Energy Adaptive Clustering Hierarchy (LEACH), focus on EE routing but often neglect the optimization of CH selection [3], which can lead to suboptimal performance and reduced NL.

Dynamic clustering methods offer flexibility by forming clusters based on real-time data and network conditions, as opposed to static clustering, which predefines clusters without regard to network dynamics. The process of selecting optimal CH have a vital part in balancing EC across the network. Effective CH selection ensures that high-energy nodes take on leadership roles [4], improving network robustness, fault tolerance, and load distribution. However, random or suboptimal CH selection can cause premature energy depletion, node failure, and network partitioning. Consequently, there is a need for advanced optimization techniques to improve both CH selection and multipath routing.

Metaheuristic optimization (MHO) algorithms, including Particle Swarm Optimization (PSO), have been applied to address these challenges by optimizing CH selection and routing paths [5]. These algorithms explore multiple solution spaces to achieve a balance between exploration (searching for new solutions) and exploitation (refining existing solutions). However, existing MHO approaches often fail to adapt dynamically to changes in network conditions, which limits their efficiency in real-world WSN deployments. Proximal Policy Optimization (PPO), a reinforcement learning-based method, can overcome these limitations by dynamically adjusting routing decisions based on real-time Quality of Service (QoS) metrics such as delay, network lifetime [6], and energy consumption.

This study introduces the Adaptive Reinforcement Learning-Based Data Aggregation and Routing Optimization (ARL-DARO) algorithm, a novel approach that leverages Reinforcement Learning (RL) to optimize both data aggregation and routing decisions dynamically. ARL-DARO enables sensor nodes to act as RL agents, learning to make optimal decisions over time by adjusting their routing and aggregation strategies based on real-time network conditions such as residual energy, traffic load, and link quality [7]. By integrating RL, the algorithm adapts to varying network conditions, improving energy efficiency and extending NL.

Unlike traditional algorithms that treat routing and data aggregation separately, ARL-DARO simultaneously optimizes these processes, allowing for more efficient energy usage and improved communication reliability. The algorithm dynamically selects CHs and optimal routes by considering Quality of Service (QoS) metrics [8], ensuring that nodes with higher energy and better link quality are chosen, thus minimizing energy depletion and improving network robustness [9]. Through simulation, ARL-DARO demonstrates significant improvements in energy consumption, routing reliability, and network lifetime compared to traditional CBR methods. These results make ARL-DARO an ideal solution for enhancing the performance and scalability of WSNs in real-world deployments.

2. LITERATURE REVIEW

A Multiple Weight LEACH (MW-LEACH) had been suggested by El Khediri et al. [10]. In MW-LEACH, the optimal number of member nodes (MN), the Residual energy (RE), and the distances among the CH are used for CHS. By choosing nodes from the first set according to the high RE nearer the density centre, an initial set of CH candidates is created. Subsequently, the candidates set out in different directions to collect data from their supporters and transfer it back to the Base Station (BS).

The suggested method is less complicated in terms of message and time. In addition, it provides a longer NL and is quick. Additionally, it offers the right amount of (FT) Fault Tolerance. When evaluating the experimental simulation, based on performance criteria including throughput, EC, packet delivery, NL, and latency, their solution performs better than state-of-the-art protocols.

An Extended Power Efficient Gathering in Sensor Information Systems (E-PEGASIS) protocol was created by Sadhana et al. [11] and is based on the PEGASIS protocol, allowing for more EE Data Transmission (DT). In this suggested approach, the BS is fixed as the radio range value of the outermost node, and the average distance between the SN is taken into consideration as the condition for chaining. It then establishes a radio link with every relevant node in the range. To proceed with the chaining process and enhance the DT efficiency among the BS and the SN, the connected node initially measures its distance from the subsequent nearest end node. In comparison with the LEACH and PEGASIS protocols, the simulation of the suggested study shows that the NL is increased.

Samraj and Shobana [12] proposed an innovative Trust Route method for multipath routing optimization, which emphasizes careful evaluation of trust factors and QoS metrics. Routing path optimization is essential in modern networking settings to sustain high performance and dependability. In order to choose routing pathways that optimize network security and efficiency, the suggested optimization method incorporates QoS metrics, such as delays, energy consumptions, NL, and distances, together with trust considerations. By merging features from the Average and Subtraction-Based Optimiser (ASBO) algorithms, the approach produces improved routing outcomes.

The approach’s usefulness and robustness in improving multipath routing (MR) methods are proved through thorough simulations and tests. The outcomes demonstrate how flexible the algorithm is to changing network conditions while still achieving the best possible performance in terms of reliability and QoS metrics.

Shobana and Samraj [13] introduced a proposed a Modified Golden Eagle Optimization with Stooping Technique (MGEO) was created utilizing the Trust Enabled Data Gathering Technique (TEDGTMGEO). A method inspired by nature is proposed by TEDGTMGEO for the selection of secure CH in MGEO. Considering the significance of node vitality and dependability. The fitness algorithm takes the node’s trust value and remaining energy into account while choosing CH. In order to ensure network dependability and prolong network lifespan, quality of service variables like energy usage and data transmission rate are considered. The efficiency of the algorithm is assessed for each iteration based on the overall energy consumption. The outcomes of the experiment indicate that compared to previous research in the literature, the proposed technique chooses more secure nodes, has a longer average network lifespan, and uses less energy.

To extend the NL, Saleem and Alabady [14] designed an EE Multipath Clustering with Load Balancing (EEMCL). The recommended protocol, which separates the network into layers of clusters, would be implemented using Multi-Hop (MH). To transmit sensor sensing data to the sink, the primary CH in each layer collaborate with the CH in the higher levels. Simulation findings compare the suggested method with SEP, SEP-E, and SEPFL methods and demonstrate improvements in network stability, EC, and NL. The last dead node is at round 5833 for SEPFL, 4027 for SEP-E, and 2325 for SEP, per the recommended protocol.

In order to balance data accuracy and DT reliability, Dan et al. [15] introduced the EE MR Algorithm(EMRA) . A multi-objective (MO) programming issue is developed from the multipath routing problem with the goal of maximizing power consumption and dependability within data accuracy limitations. An adaptive artificial immune algorithm is used to solve the MO programming problem. It improves the immune operation, antibody incentive calculation, and antibody initialization methods, particularly for the multipath routing scheme. Comparing the EMRAR algorithm to other algorithms, simulation results demonstrate that it efficiently strikes a compromise between energy savings and data quality and DT dependability.

For separating the entire sensor Net into clusters, Akbar et al. [16] presented a region-based EE MR (REMR) technique, with many candidates ideally representing each cluster. Routing packets via different clusters is done by the cluster representatives, or CRs. When routing, each route’s energy requirements are considered, and the way with the lowest energy requirements is chosen. In a similar vein, packet routing considers PDR, increased throughput, and E2E delay.

Sathiya and Nandhakumar [17] introduce a multi-path routing technique public and private key cryptography highlighting on conserving energy and secure DT in a WSN. The optimal multipath routing technology addresses issues with power consumption, dependable data transfer, and security. The main focus of this research is EE routing for data security. This method sets up three phases for node-to-node communication: path discovery, data transmission, and path management with data security. This technique secures the communication by using public cryptography, which is initiated by the source node. The security of the packets and the reduction of packet loss during transmission are ensured by this authentication and authorization. The results of the installation demonstrate reduced packet loss and increased energy usage while maximizing packet delivery ratio.

A fuzzy-based multipath clustering technique has been suggested by Patra et al. [18] and shows evidence of both static and dynamic clustered formation. The targeted region starts the clustering process when the SN are ready to begin the DT mechanism. The suggested method operates in two stages: a) MO agent-based multipath routing protocols (RP) for shortest route path discovery; and b) fuzzy CHS. The key component is the progress made in cluster development and selection. The detrimental effects of network collision and energy exhaustion have decreased because to a well-organized sensor ecosystem. Performance metrics including the PDR, communication overhead (CO), and EC are examined when simulating the specified protocol with the computer language NS2. The outcomes demonstrate that the AODV (Ad-hoc on-demand distance vector routing) protocol is not as effective as the suggested technique.

Rangappa and Dyamanna [19] proposed an Adaptive Hybrid Cuckoo Search (AHCS), the cuckoo search algorithm (CSA) population evolution strategy and Lévy Flight (LF) method are enhanced by AHCS, which also adds a mutation operation operator. Inspired by the Grey Wolf Optimisation (GWO) algorithm's concept of position update, this study presents the inertia weight w in the LF approach of the CSA. For parameters α and β, it offers new dynamic adjustment techniques. Evaluation of the suggested method considers response time, PDR, EC, E2E delay, and dead time. Additionally, the Yellow Saddle Goatfish Algorithm (YSGA), Evolutionary Multipath EER protocol (EMEER), and conventional algorithms are used to compare the performance of AHCS-GWO. For 200 nodes, the AHCS-GWO method's RE is nodes is 0.57J, it is superior when compared to the current approaches.

3. PROPOSED METHODOLOGY

A novel multi-hop routing (MR) protocol called ARL-DARO is presented in this paper. The protocol focuses on optimizing the routing process by introducing an efficient DA mechanism. When a SN collects combined data from neighboring nodes, it aggregates both the received and locally observed data using a novel aggregation technique. The protocol ensures the DA is transmitted to the next node via the shortest MR path, promoting efficient data transmission across the network. ARL-DARO improves the overall routing process by strategically selecting the shortest and most efficient paths, reducing delays and enhancing communication reliability. Based on experimental results, the ARL-DARO protocol demonstrates superior performance compared to other methods, particularly in its ability to improve routing efficiency and data aggregation. The Fig.1 provides a visual overview of the proposed protocol's procedure, highlighting its operational flow and effectiveness.

3.1 NETWORK MODEL

In this study, the network model assumes a field equipped with various sensors, including temperature, humidity, and photo sensors, each with unique sensing intervals based on specific operational requirements. Each SN has buffered that store both its observed data and data received from one-hop neighbor nodes. To manage the data effectively, sensor nodes maintain multiple buffers, each dedicated to a specific sensor type. Due to the significant correlations between neighboring nodes' same-sensor data, nodes can perform data aggregation before transmission. As depicted in Fig.2 [20], this aggregation process reduces the data redundancy by combining data from sensors of the same type at each node, optimizing network efficiency and reducing the overall communication load. This model enhances energy efficiency and improves data transmission by leveraging the spatial correlation among neighboring nodes, which is crucial for extending the network's lifetime in resource-constrained environments.

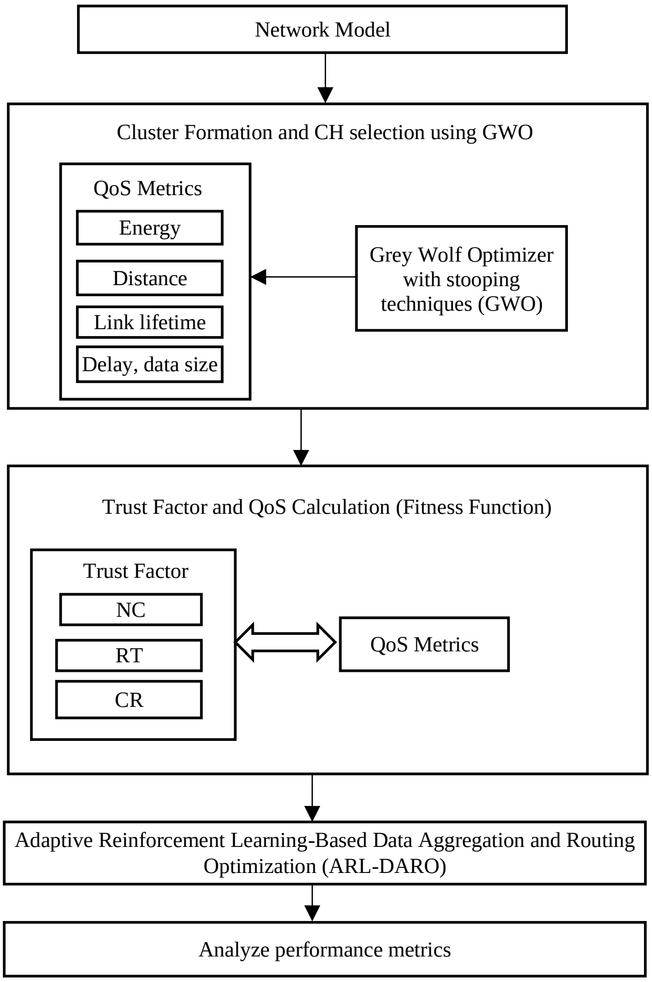

Fig.1. Overall Workflow of Proposed Framework

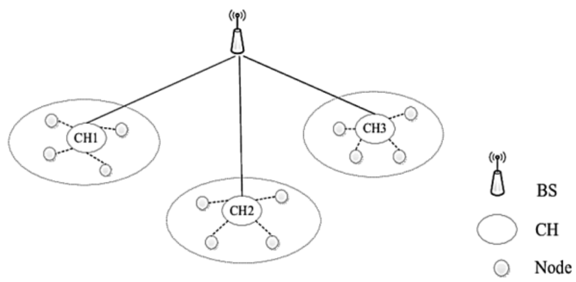

Fig.2. WSN Network Model

3.2 CLUSTER FORMATION USING GWO

For cluster formation in WSN, Adaptive Control of Exploration and Exploitation (ACE) exploration capabilities of GWO method, particularly in situations when nodes are distributed in 2-D plane [21-22]. The following are the steps:

3.2.1 Initialization: (Sensor Node and Wolf Positioning):

- **Sensor Node Deployment**: Randomly deploy sensor nodes across a 2-D plane, with each node having **(x, y)** coordinates representing its position.
- **Wolf Population Initialization**: Randomly assign wolves (candidate CHs) within the 2-D space. Each wolf represents

a potential Cluster Head, and their initial positions correspond to sensor node locations.

3.2.2 Fitness Function:

Define the FF based on QoS metrics, like:

- RE of the SN
- Distance of nodes to the sink
- Delay of the network

3.2.3 ACE Strategy - Trade-off Control:

The ACE mechanism is designed to dynamically adjust the balance between:

- Exploration: Finding new CH candidates (searching globally for better solutions)
- Exploitation: Refining the current best CH candidates (locally improving solutions)

3.2.4 Dynamic Adjustment:

ACE controls GWO parameters to $\vec{A}$ and $\vec{C}$ favor exploration in early stages and focus on exploitation as the algorithm converges in Eq.(1):

$$\vec{X}(t+1) = \vec{X}_{\alpha}(t) - \vec{A} \cdot \left| \vec{C} \cdot \vec{X}_{\alpha}(t) - \vec{X}(t) \right| \tag{1}$$

where:

$\vec{X}_{\alpha}(t)$ - position of the alpha wolf (best CH)

The coefficient vectors are $\vec{A}$ and $\vec{C}$ that adapt during exploration and exploitation.

$\vec{X}(t)$ - current position of a wolf (SN).

The ACE mechanism modifies $\vec{A}$ and $\vec{C}$ dynamically to enhance the search.

3.2.5 Direction and Parameters:

The wolves update their positions based on the alpha, beta, and delta wolves (the best CH candidates). During this process, ACE dynamically adjusts the search behavior by controlling how aggressively the wolves explore new areas or exploit the current best solutions in Eq.(2) and Eq.(3),

$$\vec{A} = 2 \cdot \vec{a} \cdot \vec{r}_1 - \vec{a} \tag{2}$$

$$\vec{C} = 2 \cdot \vec{r}_2 \tag{3}$$

To balance exploration and exploitation, $\vec{a}$ drops linearly from 2 to 0 across iterations. For stochastic exploration, the random vectors $\vec{r}_1$ and $\vec{r}_2$ are in the interval [0, 1].

3.3 LOCAL SEARCH AND CLUSTER FORMATION

Once wolves have updated their positions, the local search focuses on refining the positions in regions where node density is higher.

- **Local Refinement**: As wolves update their positions, local search mechanisms refine their positions, particularly in dense node regions. This helps in finding more suitable CHs near clustered nodes.
- **Cluster Assignment**: Once the best CH candidates (alpha, beta, delta) are selected, other sensor nodes are assigned to these CHs based on their proximity. Non-CH nodes become members of the nearest CH cluster.

3.3.1 QoS-Aware CH Selection: Fitness Function:

The fitness function F_{CH} evaluates in Eq.(4), each candidate CH based on multiple QoS metrics:

$$\begin{aligned} F_{CH} &= \omega_1 \cdot \frac{1}{\text{Residual Energy}} + \omega_2 \cdot \text{Distance} \\ &+ \omega_3 \cdot \frac{1}{\text{Link Lifetime}} + \omega_4 \cdot \text{Delay} + \omega_5 \cdot \frac{1}{\text{Data Size}} \end{aligned} \tag{4}$$

where:

w_1, w_2, w_3, w_4, w_5 weights assigned to each QoS metric.

3.3.2 Dynamic Adaptation:

Real-time network conditions, such as node failures or EC, allow ACE to adjust the optimal balance between exploration and exploitation.

- **Evaluation and Validation**: By evaluations and validations with different clustering methods, attempts to analyze how well the modified GWO procedure performs with Adaptive Control of Exploration and Exploitation. Adaptive Control of Exploration and Exploitation can be added to the GWO algorithm to enhance its search capability in large-scale or varied deployment scenarios, and WSN effectively assisted cluster formation. To determine the best path and cluster formation, DA in WSN can be performed using the objective function OF(x).

Here is how it is defined as Eq.(5),

$$\begin{aligned} OF(X) &= SF \cdot F_1 + (1-SF) \cdot F_2 + (2-SF) \cdot F_3 \\ &+ (3-SF) \cdot F_4 + (4-SF) \cdot F_5 \end{aligned} \tag{5}$$

where

SF (Scaling Factor) adjusts the importance of different QoS metrics.

F_1, F_2, F_3, F_4, F_5 represent QoS metrics such as energy efficiency, delay, link lifetime, etc.

The scaling factor (SF) has values between 0 and 1. Delays, EC, link lifetime, and distances are some of the variables that affect cluster formation in WSNs.

- **Delay (F_1):** The sum of the transmission time, propagation delay, processing delay, and queuing delay is found in Eq.(6).

$$F_1 = \text{Transmission time} + \text{Propagation delay} + \text{Processing delay} + \text{Queuing delay} \tag{6}$$

- **Energy Consumption (F_2):** EC is essential in WSN since SN's battery power is constrained. The following factors will be employed for EC computation, they are: Total energy consumed in transmission, reception, processing, and idle listening. The following Eq.(7) can be employed to calculate EC.

$$F_2 = \text{TransmissionEnergy} + \text{ReceptionEnergy} + \text{ProcessingEnergy} + \text{IdleListeningEnergy} \tag{7}$$

- **Link Lifetime (F_3):** The link lifespan is the amount of time that a communication link between two SN is operational.

EC, transmission distance, and data rate are a few of the variables that influence it. The link lifetime formula is given by Eq.(8).

$$F_3 = \text{RemainingEnergy}/\text{EnergyConsumptionRate} \tag{8}$$

- **Distance(F_4):** Transmission ranges and EC are affected by the distances among SN. Path loss, interference, and signal strength are a few examples of issues that can affect a WSN's distance formula. One might use the path loss model of the network to create a simple distance formula in Eq.(9).

$$F_4 = f(\text{Path Loss Model,Signal Strength,Interference}) \tag{9}$$

- **Data Size (F_5):** The size of the data packets being sent between nodes, which affects energy consumption and transmission time.

3.4 TRUST AWARE (TA) MODEL

A Trust-Aware (TA) framework that integrates both Quality of Service (QoS) metrics and trust factors is presented to determine optimal communication pathways in WSNs. This approach significantly lessens the negative effects of malicious nodes, weak links, and external disturbances on the network's overall efficiency.

The TA architecture enhances data aggregation and routing performance by selecting only dependable and secure communication pathways by utilizing trust information. Trust elements are essential in WSNs to ensure that communication pathways are reliable and secure [23]. The trustworthiness of each node participating in routing is assessed using a number of factors, including cooperation among nodes, trustworthiness recently, and communication reliability.

The dependability of previous communications, the dependability of path nodes, and real-time behavior are all integrated to compute the Trust Factor (TF) for each path. This trust factor is computed using three primary components: Node Cooperation (NC), Recent Trust (RT), and Communication Reliability (CR). The trust factor TF_i for each path i is calculated by combining these components as follows the Eq.(10),

$$TF_i = \sum_{j=1}^{m_i} \left(w_1 \cdot NC_{ij} + w_2 \cdot RT_{ij} + w_3 \cdot CR_{ij} \right) \tag{10}$$

where, m_i represents the number of nodes in path i, NC_{ij} denotes the cooperation of node j in path i, RT_{ij} represents the recent trustworthiness of node j, and CR_{ij} measures the communication reliability of node j. The weighting factors w_1, w_2, w_3 reflect the relative importance of each trust component, according to the particular needs of the WSN application, it is adjusted. These parameters are crucial for dynamically responding to network conditions, ensuring secure and efficient routing decisions.

- **Node Cooperation (NC):** NC measures how effectively node j in path i cooperates with other nodes by forwarding data packets and avoiding selfish behavior. It is crucial for ensuring that nodes do not drop packets or disrupt data flow.
- **Recent Trust (RT):** RT evaluates the trustworthiness of node j based on its recent behavior. A time-decay function is typically applied to prioritize recent interactions, allowing the system to adapt quickly if a node's reliability suddenly changes.
- **CR:** CR measures the performance rate of communication through node j, including packet delivery ratios and response times. A higher CR value indicates that the node has a stable and reliable communication link.

3.5 ARL-DARO ALGORITHM

The AARL-DARO algorithm is designed to efficiently manage DA and routing decisions in WSN. Its core objective is to enhance energy efficiency, improve routing reliability, and extend NL through an adaptive mechanism driven by Reinforcement Learning (RL) [24]-[25]. ARL-DARO dynamically optimizes data transmission paths and aggregation processes based on real-time network conditions, such as node energy, traffic load, and link quality.

3.5.1 ARL-DARO Algorithm:

In ARL-DARO, each sensor node acts as an RL agent that learns to make optimal decisions over time. The agent observes the network environment, selects actions (routing and aggregation decisions), and receives rewards based on the quality of its decisions [26]. The algorithm seeks to maximize long-term rewards by adapting to dynamic network conditions, such as varying energy levels or link quality.

The basic components of the RL framework in ARL-DARO include:

- **State (*S*)**: The current state of the network, including parameters like RE, link quality, queue size, and hop count.
- **Action (*A*)**: The possible actions a node can take, such as choosing the next-hop node for routing or adjusting the data aggregation level.
- **Reward (*R*)**: A feedback signal that reflects the quality of the action based on energy consumption, delay, and transmission success.
- **Policy (π)**: The strategy that maps states to actions, guiding each node in making decisions that maximize cumulative rewards over time.

3.5.2 State and Action Representation in ARL-DARO:

The state (S) of each node includes key metrics that represent the current network conditions. For ARL-DARO, these include:

- Residual energy (S): The available energy at the node, which is crucial for optimizing energy efficiency.
- Queue length (Q): The amount of data waiting to be aggregated or transmitted.
- Link quality (L_q): The reliability and strength of communication links between nodes.
- Hop count (HC): HC to the sink (SnK) node, indicating proximity to the data destination.

The actions (A) correspond to the routing decisions (choosing the next-hop node) and DA strategies. The node must decide:

- Which neighbour node should receive the DA.
- Whether to perform data aggregation or transmit the raw data directly, depending on network conditions.

3.5.3 Reward Function in ARL-DARO

The reward function ($\mathcal{R}$) is a critical component of ARL-DARO as it guides the RL agent towards making optimal decisions. The $\mathcal{R}$ is designed to balance EE, delay, and reliability in DT [27]. The reward R for an action is defined as a combination of factors in Eq.(11):

$$R = \omega_1 \cdot (-E_c) + \omega_2 \cdot (-D) + \omega_3 \cdot P_s + \omega_4 \cdot L_q \quad (11)$$

where:

E_c is the energy consumed in the action.

D is the delay introduced in data transmission.

P_s is the packet success rate, reflecting the reliability of the transmission.

L_q is the quality of the communication link.

$\omega_1,\omega_2,\omega_3,\omega_4$ are weight factors that control the influence of each metric based on network requirements.

By rewarding energy-efficient routing and penalizing excessive delays or packet losses, the agent learns to select actions that optimize overall network performance.

3.5.4 Adaptive Routing and Data Aggregation:

- **Routing Optimization**: ARL-DARO enables nodes to make adaptive routing decisions based on real-time network feedback. The RL agent selects the next-hop node based on metrics such as RE, link quality, and proximity to the sink node. This ensures that the network adapts to changing conditions like node energy depletion or link quality degradation.

The Eq.(12) represent the routing decision at time step n is based on a combination of energy and distance metrics:

$$R_E = \frac{E^r_{s'}(n)}{E^r_{s'}(0)} - \left(\frac{d_{(s-s')}}{d_{\max}}\right)^{\beta} \quad (12)$$

where:

RE of the selected neighbor node s' is denoted as $E^r_{s'}(n)$.

the distance between the current node and the next-hop node s' is denoted as $d_{(s-s')}$.

$d_{\max}$ is the maximum communication range.

β is the path loss exponent.

This equation ensures that nodes with higher residual energy and shorter transmission distances are prioritized, extending network lifetime and improving routing efficiency.

3.5.5 Data Aggregation Optimization:

Data aggregation is a key technique for reducing communication overhead and saving energy. In ARL-DARO, each node can DA from multiple SN prior in transmitting it [28]. The agent decides the level of aggregation based on the queue state and network conditions. The queue state $Q^t_i(n)$ for sensor type t at node i at the n^{th} time step is computed as Eq.(13):

$$Q^t_i(n) = OD^t_i(n) + \sum_{j \in N_i} AD^t_i(n) \quad (13)$$

where:

$OD^t_i(n)$ represents the locally collected data at node i.

$AD^t_i(n)$ represents the aggregated data received from neighboring nodes.

The node then aggregates the data using the aggregation function:

$$AD^t_i(n) = DA\{Q^t_i(n)\} \quad (14)$$

where Eq.(14), displays the $DA\{\}$ is the data aggregation function. This function balances between minimizing transmission costs and ensuring data fidelity.

3.5.6 Energy Consumption in ARL-DARO:

Energy consumption is a crucial factor in the performance of WSNs [29], and ARL-DARO is specifically designed to optimize energy usage in both DA and routing.

The energy consumption for transmission (TX) at node i at the n^{th} time step is given by Eq.(15):

$$E^{TX}_i(n) = \sum_{\forall t} \frac{AD^t_i(n)}{B} \left(P_{txElec} + P_{amp} \left(\frac{d_{i-n^*_i}}{d_{\max}} \right)^{\beta} \right) \quad (15)$$

where:

P_{txElec} is the power required for transmission.

P_{amp} is the amplifier power.

B is the bit rate of transmission.

$d_{i-n^*_i}$ is the distance to the next-hop node for sensor type.

Similarly, the energy consumption for reception (RX) is calculated as Eq.(16):

$$E^{RX}_i(n) = \sum_{\forall t} \frac{AD^t_i(n)}{B} \left(P_{rxElec} + AD^t_i(n) \cdot E_{decBit} \right) \quad (16)$$

where:

P_{rxElec} is the power required for receiving data.

E_{decBit} is the energy consumed for decoding the data.

Finally, Eq.(17), shows the energy consumption for data aggregation (DA) is calculated as:

$$E^{DA}_i(n) = \sum_{\forall t} Q^t_i(n) \cdot E_{aggBit} \quad (17)$$

where E_{aggBit} is the energy required to aggregate one bit of data.

3.5.7 Learning and Adaptation:

The learning process in ARL-DARO is based on reinforcement learning, where each node continuously improves its decisions over time by interacting with the network and receiving feedback in the form of rewards. The algorithm adapts to changing network conditions, such as:

- **Energy depletion**: When nodes run low on energy, the algorithm learns to avoid those nodes and reroutes traffic to nodes with more available energy.
- **Link failures**: If a communication link becomes unreliable, ARL-DARO adapts by selecting alternative, higher-quality routes.
- **Traffic load**: The algorithm balances the load across multiple nodes, avoiding congestion and improving overall network performance.

The ARL-DARO algorithm provides an effective solution for optimizing DA and routing in WSN [30]. By leveraging RL, ARL-DARO adapts to network dynamics, improving energy efficiency, reducing communication overhead, and extending network lifetime. The algorithm strikes a balance between minimizing energy consumption, ensuring timely data delivery, and maintaining high-quality communication links, making it ideal for real-world WSN deployments.

4. RESULTS AND DISCUSSIONS

The MATLAB R2019a environment's system definition is used to implement the recommended systems. Microsoft Windows 10 operates easily on an Intel(R) Core (TM) i5 CPU running at 2.80 GHz and 16 GB of RAM. In a free space network model for DT, transmitters and receivers send n bits of data over a distance of d. Based on current fitness levels, the algorithm has selected the solution with the lowest hop routing as the ideal CM for aggregation into a single message. when the inter-cluster communication phase starts and the CHs continue to transmit compressed data to the BS via their radios. During inter-cluster communication, the CMs may enter a sleep state in order to conserve energy while the CH remain up and use the radio via CSMA/CA. CSMA/CA facilitates communication within and between clusters. The BS can transmit the sensing zone to reach the simulation parameters listed in Table.1 and has access to network data.

Table.1. Simulation parameters

Parameters	Values
Dimension of Network area	200×200
Nodes count	100-500
Sink Position	(100,100), (100,50), (200,200)
Initial Energy	0.5, 2, 200 J
% of CH	10-15%
E_{elec}	50 nJpb
E_{amp}	0.0013 pJ/bitpm^2
E_{fs}	10 pJ/bitpm^2
d_0	30 m
d_{max}	100 m
HMCR	0.7
PAR	0.8
Packet Size	4000 bits

The performance study is performed using MATLAB R2019a and a range of simulation examinations. The Table.1 contains the simulation's settings. The test utilized in this include a few parameters that normally equate to radio energy: E_{elec}, E_{amp}, E_{fs}, d_0, and d_{max}. The test results are given in the following sections. In the MATLAB R2019a platform, the trial values are used to model the EE, NL, and QoS parameters.

- **PDR:** By dividing the data received by the DT. PDR is computed using Eq.(18).

$$PDR = \frac{\text{data\_received}}{\text{data\_transmitted}} \times 100 \tag{18}$$

- **Packet Loss Ratio (PLR):** The packet loss referred as data loss that exist in both transmission and reception, as it transmits source node to the destination node.

It is determined by dividing the amount of received data by the total number of nodes (500). PLR is computed in WSN by using Eq.(19).

$$PLR = 100 - \left(\frac{\text{data\_transmitted}}{\text{data\_received}} \times 100 \right) \tag{19}$$

- **E2E delay:** The E2E delay is the time it takes for data to go from one destination node to another inside the same network. Using Eq.(20) the route's E2E delay was calculated.

$$E2E\ Delay = T_{data\_receiving} + T_{data\_transmitting} \tag{20}$$

- **Throughput:** Throughput is a measure of how many packets are actually transported to the base station (BS) during a single round. The term "throughput" refers to the ratio between the number of packets received by the receiver from the transmitter and the time it takes to deliver the final packet. To calculate it, use Eq.(21),

$$\text{Throughput} = \frac{\text{Datasize} \times T_{\text{data\_transmitting}}}{\text{Response\_time}} \tag{21}$$

- **Energy consumption:** In the WSN, the node's EC significantly affects both DT and network efficiency. Inefficiencies in transmission can be blamed for the development of node interference. The EC to DT, denoted by tamp, is the DT by the amplifier to the node $t_{amp}.l^2$, EC is expressed as E_{con}. Let l^2 be the energy loss and d be the distance between the cluster nodes. To transfer a m-bit of data, the energy loss be represented by l^2. Eq.(22) is then used to compute the energy.

$$E_{tx}(m,d) = E_{con} \times m + t_{amp} \times m \times l^2 \tag{22}$$

- **NL:** One important deciding metric in WSN is NL, which is measured by monitoring the time it takes for the initial sensor energy to run out. Each SN in a traditional WSN is set up to communicate with the sink across many hops in order to transmit the data it has collected.

Table.2. Packet Delivery Ratio comparison of proposed and existing methods

No. of Nodes	Packet Delivery Ratio (%)					
	FAJIT	EDAGD	FRLEEDA	QDAEER	DDQNDA-TEAMR	ARL-DARO
100	87.11	89.69	91.61	93.75	95.95	99.1
200	85.63	87.78	90.44	92.29	94.33	98.27
300	84.22	85.92	88.38	91.26	93.67	97.12
400	82.47	84.76	86.49	89.41	92.18	96.73
500	80.16	83.41	85.28	88.14	91.71	94.08

The Table.2 displays the number of nodes and PDR's evaluation with respect to DA approaches. The assessment of Packet Delivery Ratio (PDR) in relation to Data Aggregation (DA) techniques is illustrated in the Fig.4. Initially, 100 sensors are utilized, and the number gradually increases by increments of 100 until reaching 500 sensors. This simulation reveals that as the number of nodes increases, the PDR tends to decrease across all algorithms. The algorithms under comparison—FAJIT, EDAGD, FRLEEDA, QDAEER, and DDQNDA-TEAMR—demonstrate varied levels of efficiency. At 100 sensor nodes (SN), the PDR for

FAJIT, EDAGD, FRLEEDA, and QDAEER is 87.11%, 89.69%, 91.61%, and 93.75%, respectively, while the suggested system (ARL-DARO) achieves the highest PDR at 95.95%. As the number of nodes increases to 500, all techniques experience a gradual decline in PDR due to increased congestion and collisions. However, the ARL-DARO system consistently outperforms the other algorithms, maintaining the highest PDR across all node counts. This performance highlights the superior efficiency of the suggested system in ensuring packet delivery even as network density grows, demonstrating better reliability and robustness compared to the current methods.

The Table.3 displays the number of nodes and PLR's evaluation with respect to DA approaches.

Table.3. Packet loss ratio comparison of proposed and existing methods

No. of Nodes	Packet Loss Ratio (%)					
	FAJIT	EDAGD	FRLEEDA	QDAEER	DDQNDA-TEAMR	ARL-DARO
100	12.89	10.31	8.39	6.25	4.05	2.70
200	14.37	12.22	9.56	7.71	5.67	3.16
300	15.78	14.08	11.62	8.74	6.33	4.53
400	17.53	15.24	13.51	10.59	7.82	5.19
500	19.84	16.59	14.72	11.86	8.29	6.89

The assessment of Packet Loss Ratio (PLR) in relation to Data Aggregation (DA) techniques is depicted in the Fig.5. Initially, 100 sensor nodes are used, and the number of nodes is gradually increased in increments of 100, reaching a maximum of 500. This simulation reveals that as the number of nodes increases, the PLR tends to rise across all algorithms. The algorithms under comparison—FAJIT, EDAGD, FRLEEDA, QDAEER, DDQNDA-TEAMR, and ARL-DARO—exhibit varied performance in managing packet loss. At 100 sensor nodes, PLR for FAJIT, EDAGD, FRLEEDA, and QDAEER is 12%, 10%, 9%, and 7%, respectively, while ARL-DARO achieves the lowest PLR at around 3%, indicating its superior efficiency in handling packet transmission with minimal loss. As the number of nodes increases to 500, PLR values increase significantly for most algorithms, especially FAJIT and EDAGD, where PLR reaches above 15%. In contrast, ARL-DARO continues to maintain a lower PLR, around 6%, showcasing its resilience in managing congestion and packet loss in denser networks. This performance highlights the superior ability of ARL-DARO to ensure reliable communication as network size grows, demonstrating its robustness compared to the other algorithms.

When compared to other existing systems, Table 4 shows how various data aggregation techniques assess E2E latency; the recommended solution has the lowest delay.

Table.4. E2E Delay comparison of proposed and existing methods

No. of Nodes	E2E Delay (sec)					
	FAJIT	EDAGD	FRLEEDA	QDAEER	DDQNDA-TEAMR	ARL-DARO
100	8.71	7.26	6.11	4.92	3.19	2.23
200	9.58	8.43	7.19	5.97	4.55	3.15
300	10.65	9.54	8.48	7.05	5.58	4.49
400	11.97	10.69	9.53	8.37	7.05	5.61
500	13.64	11.78	10.44	9.36	8.28	6.77

The assessment of End-to-End (E2E) delay in relation to different Data Aggregation (DA) techniques is depicted in the Table.4. Starting with 100 sensor nodes (SN), the number increases incrementally to 500 nodes. As shown, the E2E delay generally increases with the number of nodes due to higher traffic and congestion, though the performance varies among the algorithms. At 100 nodes, FAJIT shows an E2E delay of approximately 8 seconds, while ARL-DARO demonstrates superior performance with a delay of around 4 seconds. As the number of nodes increases to 500, the delay in FAJIT grows significantly, reaching over 12 seconds, while EDAGD, FRLEEDA, and QDAEER exhibit intermediate delays between 7 to 10 seconds. In contrast, ARL-DARO consistently maintains the lowest E2E delay, even with 500 nodes, staying below 6 seconds. The performance disparity emphasizes the efficiency of ARL-DARO, which not only reduces packet loss and ensures high packet delivery but also minimizes delay, making it the most suitable approach for networks with high node density. In contrast, the FAJIT algorithm consistently exhibits the highest delay, making it less efficient for real-time or delay-sensitive applications. Overall, ARL-DARO outperforms the other systems, maintaining low E2E delay across all node configurations.

The throughput comparison between the suggested and current methods is displayed in Table.5. The QDAEER system has a throughput of 0.9125, the FRLEEDA system has 0.8717, the EDAGD system has 0.8236, and the FAJIT system has 0.7952 and the DDQNDA-TEAMR system has throughput of 0.9512, The proposed system has a throughput of 1.2366 in 100 nodes.

Table.5. Throughput Comparison of Proposed and Existing Methods

No. of Nodes	Throughput (kbps)					
	FAJIT	EDAGD	FRLEEDA	QDAEER	DDQNDA-TEAMR	ARL-DARO
100	0.7952	0.8236	0.8717	0.9125	0.9512	1.2366
200	0.7467	0.7824	0.8249	0.8758	0.9165	1.1955
300	0.6921	0.7365	0.7716	0.8243	0.8694	1.0522
400	0.6546	0.6898	0.7244	0.7751	0.8153	1.111
500	0.6014	0.63214	0.6763	0.7216	0.7569	0.9712

The Table.5 illustrates the throughput performance of different systems—FAJIT, EDAGD, FRLEEDA, QDAEER, DDQNDA-TEAMR, and ARL-DARO—as the number of nodes increases. Throughput generally decreases as the network grows, likely due to increased congestion and communication overhead. However, ARL-DARO consistently outperforms other systems in terms of throughput, particularly with 100 nodes where it achieves 0.9512 kbps, while QDAEER records 0.9125 kbps, FRLEEDA 0.8717 kbps, EDAGD 0.8236 kbps, and FAJIT 0.7952 kbps. This superior performance can be attributed to ARL-DARO's effective use of Double Deep Q-Network (DDQN) aggregation, optimized routing strategies, and enhanced Cluster Head Selection (CHS), which help maintain higher throughput even as the node count increases. For instance, at 500 nodes, ARL-DARO continues to lead with the highest throughput, demonstrating its robustness in managing network scalability and efficiency compared to other systems.

The Table.6 displays the NL evaluation across DA approaches. The proposed system is found to have a greater NL than the existing algorithms.

Table.6. Network lifetime comparison of proposed and existing methods

No. of Nodes	Network Lifetime (rounds)					
	FAJIT	EDAGD	FRLEEDA	QDAEER	DDQNDA-TEAMR	ARL-DARO
100	4336	4742	5165	5448	5789	8393
200	4154	4426	4813	5125	5467	8089
300	3915	4239	4564	4892	5178	7912
400	3726	4047	4347	4648	4915	7677
500	3457	3898	4205	4451	4763	7595

The Table.7 shows the NL performance of different DA techniques—FAJIT, EDAGD, FRLEEDA, QDAEER, DDQNDA-TEAMR, and ARL-DARO—across varying node counts. The primary aim is to maximize the network's operational rounds, and ARL-DARO consistently achieves the highest lifetime across all scenarios. For instance, with 500 nodes, ARL-DARO reaches a lifetime of around 8,000 rounds, significantly surpassing QDAEER at 4,763 rounds, FRLEEDA at 4,451 rounds, EDAGD at 4,205 rounds, and FAJIT at 3,898 rounds. FAJIT consistently has the shortest lifetime, indicating less efficiency in energy management and resource allocation compared to other methods. As the number of nodes increases, the network lifetime generally decreases across all systems, but ARL-DARO's advanced energy-efficient routing and clustering mechanisms allow it to maintain a longer network lifespan. This suggests that ARL-DARO is particularly effective in handling higher node densities, offering superior performance in large-scale networks compared to the alternatives.

5. CONCLUSION

In conclusion, this research introduces the ARL-DARO algorithm to address the critical risks of EC and NL in WSN. By employing the GWO for efficient cluster formation and Cluster Head (CH) selection, ARL-DARO optimizes energy usage while ensuring reliable routing. The integration of trust factors like Node Connectivity, Residual Trust, and Cooperation Rate into QoS metrics enhances both security and route reliability. Furthermore, the application of Q-learning for adaptive DA and routing improves network adaptability, reducing data redundancy and extending the network's lifespan. Performance metrics such as energy consumption, throughput, PDR, E2E delay, and NL demonstrate that ARL-DARO significantly overtakes current methods like TEAMR and DDQNDA. Future research will focus on further enhancing the model by addressing mobility and coverage challenges, aiming to extend the overall network lifetime even further.

REFERENCES

[1] M. Faris, M.N. Mahmud, M.F.M. Salleh and A. Alnoor, "Wireless Sensor Network Security: A Recent Review based on State-of-the-Art Works", *International Journal of Engineering Business Management*, Vol. 15, pp. 1-6, 2023.

[2] A. Panchal and R.K. Singh, "Energy Efficient Hybrid Clustering and Hierarchical Routing for Wireless Sensor Networks", *Ad Hoc Networks*, Vol. 123, pp. 1-9, 2021.

[3] I.A. Ezenugu, "Low-Energy Adaptive Clustering Hierarchy Algorithm for Selection of Cluster Head in Wireless Sensor Nodes", *Journal of Multidisciplinary Engineering Science and Technology*, Vol. 7, No. 1, pp.12327-12335, 2020.

[4] M. Gheisari, M.S. Yaraziz, J.A. Alzubi, C. Fernandez-Campusano, M.R. Feylizadeh, S. Pirasteh, A.A. Abbasi, Y. Liu and C.C. Lee, "An Efficient Cluster Head Selection for Wireless Sensor Network-based Smart Agriculture Systems", *Computers and Electronics in Agriculture*, Vol. 198, pp. 1-6, 2022.

[5] K. Zaimen, L. Moalic, A. Abouaissa and L. Idoumghar, "A Comparative Study of Meta-Heuristic Algorithms for WSN Deployment Problem in Indoor Environments", *Congress on Evolutionary Computation*, pp. 1-8, 2023.

[6] R. Li, L. Xu, C. Tang, P. Wang, W. Liu, J. Gu, Z. Cai and R. Jiang, "Proximal Policy Optimization based Resource Allocation for Energy Harvesting Industrial Wireless Sensor", *Proceedings of International Conference on Wireless Communications*, pp. 1-9, 2023.

[7] A. Mohammadi-Balani, M.D. Nayeri, A. Azar and M. Taghizadeh-Yazdi, "Golden Eagle Optimizer: A Nature-Inspired Metaheuristic Algorithm", *Computers and Industrial Engineering*, Vol. 152, pp. 1-7, 2021.

[8] H. Hu, Y. Han, M. Yao and X. Song, "Trust based Secure and Energy Efficient Routing Protocol for Wireless Sensor Networks", *IEEE Access*, Vol. 10, pp.10585-10596, 2021.

[9] R. Rastogi, S. Srivastava, M.S. Manshahia and N. Kumar, "Withdrawn: A Hybrid Optimization Approach using PSO and Ant Colony in Wireless Sensor Network", *Proceedings of International Conference on Wireless Communications*, pp. 1-9, 2021.

[10] S. El Khediri, R.U. Khan, N. Nasri and A. Kachouri, "MW-LEACH: Low Energy Adaptive Clustering Hierarchy Approach for WSN", *IET Wireless Sensor Systems*, Vol. 10, No. 3, pp.126-129, 2020.

[11] S. Sadhana, E. Sivaraman and D. Daniel, "Enhanced Energy Efficient Routing for Wireless Sensor Network using Extended Power Efficient Gathering in Sensor Information Systems Protocol", *Procedia Computer Science*, Vol. 194, pp.89-101, 2021.

[12] M.J. Samraj and V. Shobana, "Trustroute: A Novel Approach for Qos-Enabled Data Gathering and Multipath Routing Optimization in Wireless Sensor Networks", *Proceedings of International Conference on Wireless Communications*, pp. 1-7, 2023.

[13] V. Shobana and J. Samraj, "A Novel Trust-Enabled Data-Gathering Technique based on Modified Golden Eagle Optimization in Wireless Sensor Network", *Proceedings of International Conference on Computing and Data Science*, pp. 1-6, 2024.

[14] M.M. Saleem and S.A. Alabady, "Energy-Efficient Multipath Clustering with Load Balancing Routing Protocol

for Wireless Multimedia Sensor Networks", *IET Wireless Sensor Systems*, Vol. 13, No. 3, pp.104-114, 2023.

[15] F. Dan, Y. Ma, W. Yin, X. Yang, F. Zhou, S. Lu and B. Ning, "An Accuracy-Aware Energy-Efficient Multipath Routing Algorithm for WSNs", *Sensors*, Vol. 24, No. 1, pp. 285-298, 2024.

[16] S. Akbar, M.M. Mehdi, M.H. Jamal, I. Raza, S.A. Hussain, J. Brenosa, J.C.M. Espinosa, A.E. Pascual Barrera and I. Ashraf, "Multipath Routing in Wireless Body Area Sensor Network for Healthcare Monitoring", *Healthcare*, Vol. 10, No. 11, pp. 2297-2302, 2022.

[17] M. Sathiya and R. Nandhakumar, "An Optimized Multipath Routing for Secure Communication of Wireless Sensor Network", *International Journal of Advanced Research and Technology*, Vol. 8, No. 3, pp.1-6, 2020.

[18] B.K. Patra, S. Mishra and S.K. Patra, "A Novel Agent-based Multipath Routing Protocol to Extend Lifetime and Enhancing Reliability of Clustered Wireless Sensor Networks", *International Journal of Computer Networks and Applications*, pp. 1-6, 2022.

[19] M.H. Rangappa and G.C. Dyamanna, "Energy-Efficient Cluster based Routing for Hybrid Wireless Sensor Networks using Adaptive Hybrid Cuckoo Search and Grey Wolf Optimization Algorithm", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 5, 2021.

[20] B.A. Begum and S.V. Nandury, "A Survey of Data Aggregation Protocols for Energy Conservation in WSN and IoT", *Wireless Communications and Mobile Computing*, No. 1, pp. 1-7, 2022.

[21] M. Otair, O.T. Ibrahim, L. Abualigah, M. Altalhi and P. Sumari, "An Enhanced Grey Wolf Optimizer-based Particle Swarm Optimizer for Intrusion Detection System in Wireless Sensor Networks", *Wireless Networks*, Vol. 28, No. 2, pp.721-744, 2022.

[22] R. Ahmadi, G. Ekbatanifard and P. Bayat, "A Modified Grey Wolf Optimizer-based Data Clustering Algorithm", *Applied Artificial Intelligence*, Vol. 35, No. 1, pp.63-79, 2021.

[23] S. Kumar and S. Chattopadhyay, "TRQP: Trust-Aware Real-Time QoS Prediction Framework using Graph-based Learning", *Proceedings of International Conference on Service-Oriented Computing*, pp. 143-152, 2022.

[24] S. Sarwar, R. Sirhindi, L. Aslam, G. Mustafa, M.M. Yousaf and S.W.U.Q. Jaffry, "Reinforcement Learning based Adaptive Duty Cycling in LR-WPANs", *IEEE Access*, Vol. 8, pp.161157-161174, 2020.

[25] E. Dvir, M. Shifrin and O. Gurewitz, "Cooperative Multi-Agent Reinforcement Learning for Data Gathering in Energy-Harvesting Wireless Sensor Networks", *Mathematics*, Vol. 12, No. 13, pp. 1-7, 2024.

[26] M. Krishnan and Y. Lim, "Reinforcement Learning-based Dynamic Routing using Mobile Sink for Data Collection in WSNs and IoT Applications", *Journal of Network and Computer Applications*, Vol. 194, pp. 1-7, 2021.

[27] W.K. Ghamry and S. Shukry, "Multi-Objective Intelligent Clustering Routing Schema for Internet of Things Enabled Wireless Sensor Networks using Deep Reinforcement Learning", *Cluster Computing*, pp.1-21, 2024.

[28] M.U. Younus, M.K. Khan, M.R. Anjum, S. Afridi, Z.A. Arain and A.A. Jamali, "Optimizing the Lifetime of Software Defined Wireless Sensor Network Via Reinforcement Learning", *IEEE Access*, Vol. 9, pp.259-272, 2020.

[29] Z. Mammeri, "Reinforcement Learning based Routing in Networks: Review and Classification of Approaches", *IEEE Access*, Vol. 7, pp.55916-55950, 2019.

[30] A. Musaddiq, T. Olsson and F. Ahlgren, "Reinforcement-Learning-based Routing and Resource Management for Internet of Things Environments: Theoretical Perspective and Challenges", *Sensors*, Vol. 23, No. 19, pp. 8263-8275, 2023.