I. J. Computer Network and Information Security, **2026, 2, 112-125**
Published Online on April 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2026.02.07

Reinforcement Learning-Based Self-Healing Routing in Fault-prone Wireless Sensor Networks

Dipti Chauhan*
Department of Artificial Intelligence & Data Science, Prestige Institute of Engineering Management & Research Indore M.P., India
E-mail: diptichauhan82@gmail.com
ORCID iD: https://orcid.org/0000-0003-1665-7587
Corresponding author

Pritika Bahad
Department of Artificial Intelligence & Data Science, Prestige Institute of Engineering Management & Research Indore M.P., India
E-mail: bahad.pritika@piemr.edu.in
ORCID iD: https://orcid.org/0000-0002-3894-4125

Jay Kumar Jain
Department of Mathematics, Bio-informatics & Computer Applications, Maulana Azad National Institute of Technology Bhopal, M.P., India
E-mail: jayjain.research@gmail.com
ORCID iD: https://orcid.org/0000-0002-9590-0006

Received: 16 November 2025; Revised: 02 January 2026; Accepted: 25 February 2026; Published: 08 April 2026

Abstract: Wireless Sensor Networks consist of energy constrained sensor nodes that monitor and transmit data to a central base station. These networks are highly susceptible to link and node failures, which further degrades performance and reduce overall network reliability. In this paper we have addresses these challenges and proposed a reinforcement learning based self-healing routing (RL-SHR) protocol, implemented in NS2 simulation environment. In the work, each node functions as an autonomous RL agent that learns optimal routing paths by interacting with the network environment and adapting to failure conditions. The protocol enables nodes to dynamically avoid unreliable paths, recover from faults, and optimize performance over time. Simulation results shows that the proposed protocol significantly outperforms traditional routing protocols such as AODV and DSR in terms of packet delivery ratio, end-to-end delay, energy consumption and network lifetime under varying failure scenarios. This work lays the groundwork for integrating learning based resilience mechanisms into next generation sensor networks.

Index Terms: Fault Tolerance, Q-Learning, Self-Healing Routing, Reinforcement Learning (RL), Wireless Sensor Networks (WSNs).

1. Introduction

The progress of pervasive and intelligent systems largely depends on Wireless Sensor Networks (WSNs). These networks consist of small, resource-constrained sensor nodes deployed over a region to monitor environmental or physical conditions and relay data to a central processing unit, typically called a sink [1]. UAVs play a role in environmental monitoring, smart farming, automating factories and power plants, surveillance on the battlefield and running hospitals. WSNs are capable of many uses yet, because they are fragile, they are commonly deployed in harsh, inaccessible, or unpredictable environments. As a result, such networks are sensitive to node failures, unreliable wireless links, and dynamic network topologies. It is still a critical research challenge to ensure networks running, data reliability, and adaptive routing in such unstable environments [2].

The recurrent failure of the components is one of the major problems associated with the deployment of sensor networks. As noted earlier, unattended sensor nodes after deployment are most frequently battery-operated, have limited processing and communicating resources, and locked capabilities [3]. Power shortage, hardware issues, physical damages, or unfavorable environmental conditions are some of the factors that cause a node to be inoperative. Failing links within

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

wireless communication can easily occur from interference, fading, and congestion. Standard protocols such as Ad hoc On-Demand Distance Vector (AODV) or Dynamic Source Routing (DSR) are not suitable to handle unpredictably recurring failures [4]. In these dynamic and fault-prone environments, the protocols rely on a reactive route discovery mechanism. Also, they do not adapt to changing network conditions and integrate past experiences, which can make performance in volatile environment less than optimal.

The concept of self-healing was developed to mitigate the weakness of WSNs. A self-healing network is one that can detect issues, remove them from the network, and automatically recover by rerouting data through functioning nodes without manual assistance [5]. This feature is very important for the preservation of integrity and continuity of operations and communications in the network. Most of the self-healing algorithms proposed by the researchers uses some form of static or deterministic recovery processes. These processes are not neither adaptable of adjusting to changes in network conditions, nor optimize performance over time. Due to this, there is a need on developing intelligent routing mechanism that uses a combination of artificial intelligence, machine learning and deep learning algorithms. The most relevant of these for allowing WSNs to adapt to their distributed and dynamic conditions is reinforcement learning (RL). This technique uses a system of rewards for autonomous agents to interact with their environment and determine the most favorable policies for action [6]. Each node in WSN can function as RL agent that learns to select the optimal, trusted and energy efficient data transmission routes. By evaluating feedback from the environment such as- delivery, energy consumption and link quality, each node updates its routing decisions without requiring centralized control. This decentralized learning process helps the network adapt to dynamic conditions, heal from failures, and ensure reliable communication while managing energy consumption in the network.

Reinforcement Learning (RL) as opposed to supervised learning does not require labeled data and is therefore appropriate for dynamic environments where the network conditions are constantly changing. RL allows nodes to learn routing strategies based on the results of their actions rather than pre-defined failure conditions [7]. This allows the network to adapt for dynamic reconfiguration of routes in response to failures. RL offers flexibility to optimize multiple performance objectives such as- energy efficiency, latency and delivery reliability, through the usage of reward function.

This area of using reinforcement learning in WSN routing has not been widely studied and particularly not through using simulation to test how well the system works in real-life networks. The majority of the work either suggests centralized learning which is difficult in distributed sensor systems or fails to include practical tests on networks with simulators such as NS2 or NS3. The lack of practical implementations and performance benchmarks creates a gap in understanding the viability of RL-based routing in real-world WSN deployments. Furthermore, few studies investigate how RL can be used explicitly to enable self-healing routing in the presence of node and link failures—an increasingly important consideration in modern WSN applications.

In this paper we address this gap by proposing a novel reinforcement learning-based self-healing routing (RL-SHR) protocol for wireless sensor networks with high failure rates. This protocol leverages Q-learning method, to allow each node in the network to learn optimal forwarding strategies by interacting with its local environment. A decentralized learning architecture is introduced in this protocol where nodes automatically adapt routing decisions based on real time feedback. This includes delivery success, energy depletion and link/node failure information. To validate the performance of proposed protocol, implementation was done using NS2 simulator. Extensive experiments were conducted under various fault scenarios including random node failure, link failure and energy based node shutdowns. The proposed RL-SHR protocol is compared against standard protocols such as AODV and DSR to assess its performance in terms of packet delivery ratio, end-to-end delay, energy efficiency and network lifetime.

The key focus area of this research is to:

- Develop a distributed reinforcement learning framework that enables individual WSN nodes to autonomously learn and adapt their routing strategies to maintain communication in the presence of faults.
- Introduce a localized fault detection approach that empowers nodes to identify link or node failures and trigger route reconfiguration through reinforcement learning updates.
- Extend the NS2 simulator with agent-based reinforcement learning capabilities to support realistic testing of dynamic and self-adaptive routing protocols.
- Perform a comprehensive simulation of the proposed protocol across various node and link failure scenarios, benchmarking against traditional protocols (e.g., AODV, DSR) using metrics such as delivery ratio, latency, and energy consumption.
- Analyze the learning process of RL agents in a dynamic environment to assess the convergence rate, adaptability, and stability of the routing strategy in the presence of persistent changes in the network.

Rest of the paper is structured as follows- Section 2 presents an overview of the existing literature on fault-tolerant routing and machine learning techniques in WSNs. Section 3 describes the system model, assumptions, and problem definition, as well as the reinforcement learning model. Section 4 presents the design of the proposed RL-SHR algorithm and its implementation. Section 5 describes the simulation environment in NS2, parameters, fault model, and simulation results. Section 6 summarizes the Results & Discussion. Section 7 concludes the paper and provides recommendations for future work.

2. Related Works

WSN networks have often faced challenges due to failures and shortages of energy, so fault-tolerant routing has gained importance in research. Traditional routing protocols including AODV and DSR are widely adopted, yet they do not have features that automatically help them inherent self-adaptability to recover from faults. Even these protocols, often fail to maintain optimal performance under dynamic conditions like random node failures or link disruptions, which are common in real-world WSN deployments. Many fault-tolerant routing protocols have been suggested to improve the strength of wireless sensor networks. In particular, LEACH-F and HEED aim to mitigate energy imbalance and node failure by organizing nodes into clusters and regularly selecting new cluster leaders. Despite of this, these protocols do not adapt well when failure occurs and are unable to learn from the current network configuration. Recent advances in AI, researchers are now investigating machine learning routing techniques that can dynamically adapt to changing network environments.

Authors [8] proposed an energy efficient data collection strategy for WSNs using double Q-learning to optimize the mobile sinks trajectory. By using two Q-tables, their approach reduces overestimation bias and achieves better convergence than standard Q-learning.

Authors [9] developed a reinforcement learning based routing protocol focused on maximizing energy efficiency in static WSNs. Each node learns to select optimal paths based on local observations of energy and link status. The proposed model outperforms the traditional methods in terms of lifetime and packet delivery ratio.

The authors in [10] investigated data-driven ML methods for self-healing in cellular wireless networks, specifically addressing fault detection, diagnosis, and self-healing. Their research focuses on cellular wireless networks rather than WSNs, although the idea of learning-assisted self-healing matches well with RL-SHR. Both methods strongly emphasize real-time resilience and intelligent fault recovery, although RL-SHR does so in a fully distributed fashion that is more suitable for resource-constrained sensor nodes.

The authors in [11] designed an energy-efficient routing protocol for software-defined wireless sensor networks using reinforcement learning. Their design combines Q-learning with a software-defined architecture to facilitate centralized control decision-making with a network-wide understanding. By separating the control and data planes, the design optimizes routing paths based on real-time energy costs and traffic dynamics.

The authors in [12] designed a deep reinforcement learning solution for improving energy efficiency and throughput in mobile wireless sensor networks (MWSNs). Their design employs DRL agents to dynamically control routing and mobility in real time to optimize network performance. Unlike traditional Q-learning, their design employs deep neural networks to effectively process complex state spaces, making it more appropriate for mobile and scalable networks.

However, very few studies have been able to incorporate RL algorithms within simulators such as NS2 or NS3. Most studies have been conducted within MATLAB platforms. Moreover, the application of RL algorithms for the purpose of explicit fault recovery and self-healing routing is still in its infancy. The capability of an RL node to identify faults, discover novel optimal routes, and adaptively modify its strategy in response to unpredictable fault patterns has yet to be explored. Very little existing work has been able to comprehensively implement a decentralized reinforcement learning protocol with inherent fault detection and self-healing mechanisms within NS2, particularly in high-failure conditions. The current literature either ignores dynamic learning during faults or fails to provide a scalable simulation platform to test real-time performance.

This highlights the originality of the current work, which aims to introduce an RL-based self-healing routing protocol particularly tailored for fault-ridden WSNs, fully simulated within NS2. Through performance analysis based on varying fault injection rates and comparisons with traditional protocols such as AODV and DSR, our research work establishes a real-world applicability benchmark for RL-integrated WSNs.

3. System Model and Problem Formulation

In this paper we have considered a WSN consisting of N homogeneous sensor nodes randomly deployed in a two-dimensional square sensing area of size A × A meters. Each node is equipped with limited energy, fixed communication range R, and can perform basic sensing, processing, and routing tasks. A static sink node is located at a predefined position, either within or outside the sensing field, and is responsible for collecting data from all sensor nodes.

Sensor nodes communicate with each other via multi-hop communication. Each node maintains a list of one-hop neighbors, and data packets are forwarded to the sink through dynamically discovered routes. The network operates in a fault-prone environment where nodes may fail due to energy depletion or environmental factors. Additionally, link quality may fluctuate due to interference or node mobility. To focus the model and simulation, the following assumptions are made:

- All nodes are initially aware of their neighbors through periodic HELLO messages.
- Nodes are static, but links can fail dynamically.
- The sink is static and has unlimited energy.
- Each node is capable of executing lightweight reinforcement learning (Q-learning).
- Nodes can detect their own energy level and packet delivery success or failure.

- Node failures are modeled randomly to simulate harsh or unpredictable conditions.

3.1. Fault Model

The network is prone to two kinds of faults: Node Failure: A node becomes inactive and stops to contribute to the network because of battery drain or physical damage. The failed nodes stop routing and receiving data packets.

Link Failure: The communication path between two nodes becomes interrupted due to interference, environmental factors, or temporary obstructions. The link failures are considered random events and can be temporary or permanent.

These failures can happen at any point in the simulation, and the protocol has to detect and recover from the failures in real-time.

3.2. Reinforcement Learning Framework

Each sensor node is an independent RL agent that learns to choose the best next-hop node that is most reliable and energy-efficient based on its experience. The learning process is modeled using the Q-learning algorithm, which is a popular model-free RL algorithm.

a) State Space (S)

A "state" is the local network condition perceived by a node at the time of making a decision about packet forwarding. Each node's state is defined based on local perception and network context. A state can be represented as a vector or tuple—

$$s = (\text{ResidualEnergy}, \text{LinkQuality}, \text{NeighborBufferLevel})$$

for example- At time t, node ni perceives:
Residual energy = 0.75 (normalized between 0 and 1)
Link quality to neighbor nj = 0.85 (based on signal strength or past ACKs)
Queue buffer occupancy at nj = 0.3 (indicating low congestion)
So, the state could be: s = (0.75, 0.85, 0.3)

b) Action Space (A_t)

An action corresponds to selecting a neighbouring node as the next hop. Its currently reachable neighbors define the action space at each node.

$$A_t = \{n1, n2, \ldots . nk\}, \text{where } k \leq \max \text{neighbors} \tag{1}$$

Example: If node ni has 3 neighbors: {n3, n5, n7}, then: a ∈ {n3, n5, n7}
The agent must choose one of these nodes as the next hop, based on the Q-values associated with the current state and each action.

c) Reward Function (R)

A reward is assigned based on the success or failure of packet transmission and energy metrics. The reward function encourages successful delivery, energy balance, and avoidance of failed nodes.

Example reward:

$$R = \begin{cases} +10 \ \ if\ packet\ delivered\ successfully\ to\ sink \\ -5 \ \ if\ link\ failure\ occurs \\ -2\ if\ neighbor\ is\ unresponive \\ -1\ if\ low\ residual\ energy \end{cases} \tag{2}$$

d) Q-Value Update

Each node updates its Q-values based on the feedback using the standard Q-learning update rule.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[\, R_t + \gamma \max Q\,(s_{t+1}, a') - Q(s_t, a_t)] \tag{3}$$

Where: α is the learning rate,
γ is the discount factor,
Rt is the immediate reward,
(Q(st, at) is the Q-value for state-action pair.

3.3. Self-Healing Mechanism

The protocol integrates a local fault detection mechanism that monitors- Transmission ACK timeouts, Neighbor table refresh status, Residual energy threshold crossing.

When a fault is detected e.g. link failure, neighbor inactivity, the node triggers RL update cycle to re-evaluate Q-values and select an alternate next hop. This ensures the routing decision adapts in real time, making the network self-healing and resilent.

4. Proposed Protocol Design

The proposed protocol RL-SHR integrates reinforcement learning into the routing mechanism of WSNs, enabling sensor nodes to select optimal forwarding paths even in the presence of dynamic node and link failures. The protocol emphasizes self healing, energy efficiency, and real time learning, making it suitbale for unsatble environments.

RL-SHR is a fully distributed protocol, where each node functions as an autonomous RL agent. Nodes collect local environment information, update their routing policies based on Q-learning, and route packets based on the best neighbor available. In case of failure detection, such as unresponsive nodes, link failures, or energy depletion, nodes independently trigger a learning cycle to reconfigure their routing decisions. The protocol is composed of three main modules:

4.1 Neighbor Monitoring and Fault Detection Module
4.2 Q-Learning-Based Decision Engine
4.3 Self-Healing Forwarding Module

4.1. Neighbor Monitoring and Fault Detection Module

Periodically each node sends and receives HELLO packets to maintain an up-to-date neighbor table. The node uses three indicators to detect faults

ACK Timeout: Absence of acknowledgment signals link failure.
Energy Monitoring: If a neighbor's energy falls below a threshold, it is marked unreliable.
Missing Beacons: Failure to receive periodic HELLO messages implies node failure.
This detection causes the routing layer to query the Q-values of alternative routes.

4.2. Q-Learning-Based Decision Engine

The learning engine at each node is responsible for maintaining a Q-table that holds Q-values for state-action pairs. The engine updates decisions based on interactions with the environment using the Q-learning update rule.

a) State Representation

The state is represented by a set of local parameters:
Residual Energy (Low/Medium/High)
Link Quality Index (Good/Fair/Poor)
Neighbor Count (Sparse/Dense)
Recent Packet Success Rate
States are discretized to reduce complexity.

b) Action Selection

An action is defined as the selection of a neighbor node for the next hop. The action selection is based on an ε-greedy policy:

With probability ε, select a random action (exploration).
With probability $1-\varepsilon$, select the action with the highest Q-value (exploitation).

c) Reward Calculation

The reward function in RL-SHR is formulated to encourage routing towards reliability, energy efficiency, and network sustainability. Based on the results of each attempt to forward packets, nodes are provided with a reward in the form of a numerical value that updates their Q-values. Table 1 illustrates the Reward Mapping criteria for Routing Events.

This reward system incentivizes behavior that supports packet reliability and routing load, while discouraging failure and energy-inefficient routing choices. This guarantees that the Q-learning agents will move towards optimal and failure-free routing solutions.

Table 1. Reward mapping for routing events

Event	Reward
Successful delivery	10
Link failure	-5
Neighbor energy low	-2
Packet dropped	-10
Balanced load usage	3

d) Q-Value Update

Following every attempt to send a packet, the sensor node will receive feedback from the network in the form of a success message (ACK), a timeout, or energy and link quality information. Based on this feedback, the node assigns a reward and updates its Q-table using the Q-learning update rule:

$$Q(s,a) \leftarrow Q(s,a_ + \alpha[r + \gamma \max Q(s', a') - Q(s,a)] \quad (4)$$

Where:
Q(s,a) is the current Q-value for taking action
α: learning rate (e.g., 0.1), controlling how much new information overrides old values
γ: discount factor (e.g., 0.9) determining the importance of future rewards
r: immediate reward received after taking action a
s′ is the next state observed after taking the action
max Q(s', a') = is the best predicted reward from the new states

e) Parameter Selection and Justification

The selection of parameters for RL-SHR was done with the aim of achieving a balance between learning stability and adaptability. The learning rate (α = 0.1) was set to facilitate a stable update of the Q-values. A higher learning rate was avoided since it may lead to oscillations in routing. The discount factor (γ = 0.9) gives more importance to long-term performance, which is based on the rewards of energy conservation and network lifetime extension. This is more important in WSNs, where long-term performance is essential. The exploration rate (ε = 0.2 with decay) was set to ensure that nodes occasionally explore other neighbors. This improves adaptability in dynamic fault conditions. The decay rate ensures that the network adapts from exploration in the early stages of learning to exploitation of stable energy-efficient routes after convergence. These parameters have been used in previous reinforcement learning applications in WSNs and were tested for their efficacy in preliminary simulations.

4.3. Self-Healing Forwarding Module

This module implements the routing action based on the Q-values. During the forwarding of a packet:
The node chooses the neighbor with the maximum Q-value from the set of reliable links.
If all neighbors are unreliable or unresponsive, a local route discovery is performed using Q-values and random backoff.
The node does not communicate with nodes previously identified as faulty using a blacklist for a limited time.

4.4. Protocol Workflow & Algorithm

Initialization: HELLO messages are broadcasted, and neighbor tables and Q-tables are initialized.
Data Generation: Data packets are generated by source nodes to transmit to the sink.
Routing Decision: Nodes decide the next hop based on the current state and Q-values.
Fault Detection: When a node or link fails, the detection module updates the RL.
Q-Update: The node receives a reward (positive or negative) and updates the Q-table.
Learning: Nodes learn optimal routes that are reliable and energy-efficient over time.
Repeat: The process continues for each transmission, enabling dynamic adaptation.

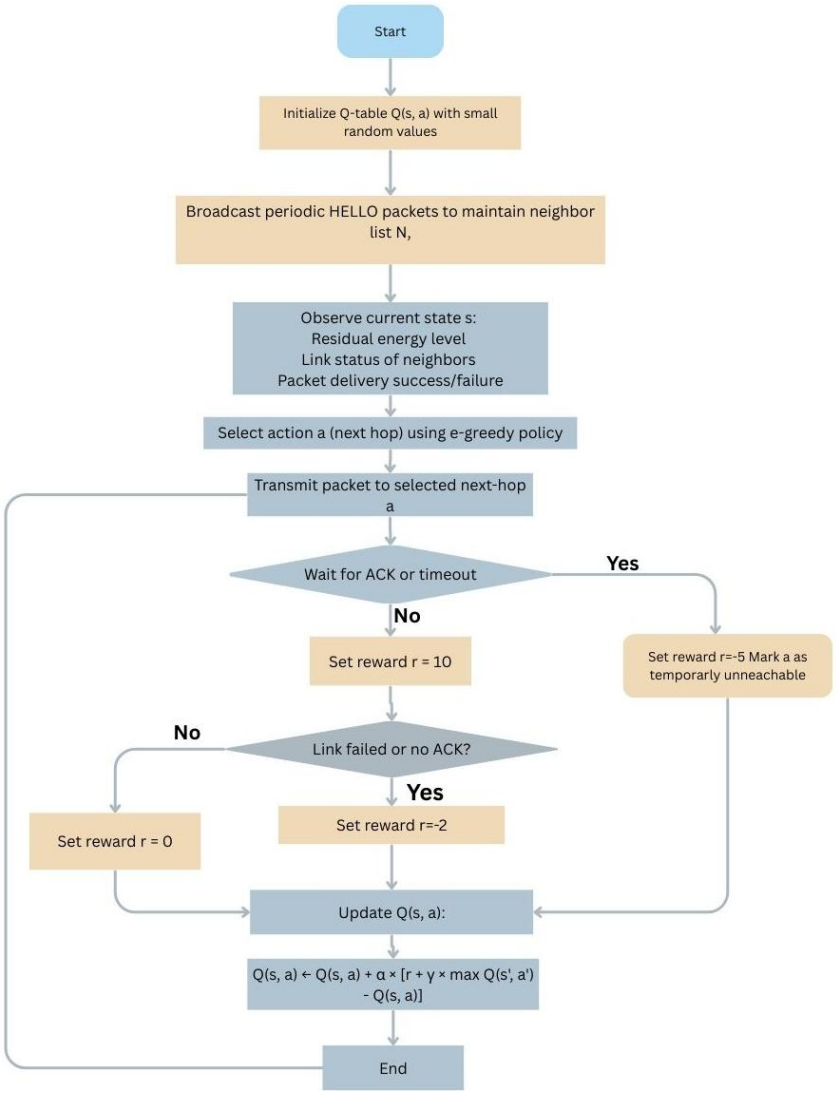


Fig.1. Flow chart- RL-SHR

Algorithm: RL-SHR

Input: Neighbor list Ni
Initial Q-table (s,a)
Energy threshold Eth
Learning rate α, discount factor γ, exploration rate ϵ
Output: Next-hop decision for data forwarding
Updated Q-table based on feedback
Initialize Q-table Q(s, a) with small random values
Broadcast periodic HELLO packets to maintain neighbor list N_i
while Node is active do
Observe current state s:
- Residual energy level
- Link status of neighbors
- Packet delivery success/failure
Select action a (next hop) using ε-greedy policy:
if rand() < ε then
Choose a random neighbor from N_i
else
Choose a = argmax Q(s, a) from N_i
end if
Transmit packet to selected next-hop a
Wait for ACK or timeout
if ACK received then

```
    Set reward r = +10
  else if link failed or no ACK then
    Set reward r = -5
    Mark a as temporarily unreachable
  else if neighbor energy < E_th then
    Set reward r = -2
  else
    Set reward r = 0
  end if
  Observe new state s'
  Update Q(s, a):
    Q(s, a) ← Q(s, a) + α × [r + γ × max Q(s', a') - Q(s, a)]
  Update state ← s'
  if all neighbors failed then
    Broadcast local route discovery request
  end if
end while
```

Flow diagram that demonstrates the algorithm is depicted in fig 1.

We employ a Q-learning framework at each node, where the agent (node) learns to select the optimal next hop based on feedback in the form of routing rewards. In our implementation, each state is modeled as a tuple of three parameters: residual energy of the node, link quality to each neighbor, and the neighbor's buffer level. Actions correspond to the selection of one of the available neighboring nodes for packet forwarding. For instance, a node with neighbors n3, n5, and n7 in a state (0.75, 0.85, 0.3) will evaluate Q-values for each neighbor and select the best candidate based on the ε-greedy policy.

5. Simulation Setup and Parameters

To validate the performance and effectiveness of the proposed RL-SHR protocol, extensive simulations were conducted using the Network Simulator 2 (NS2) platform. The simulations aimed to evaluate the protocol's adaptability, fault tolerance, and efficiency under dynamic and failure-prone wireless sensor network conditions. The simulations were carried out using NS2 version 2.35, with customized support for reinforcement learning modules integrated into the routing layer. The topology and communication patterns were graphically represented using Network Animator (NAM). TCL scripting with appropriate extensions in C++ was used to implement traffic models, fault simulation, and energy dynamics.

Table 2. Simulation parameters

Parameter	Value / Description
Simulator	NS2 v2.35
Number of sensor nodes	50, 75, 100 (varied to test scalability)
Deployment area	500m × 500m
Simulation time	1000 seconds
Traffic model	Constant Bit Rate (CBR) over UDP
Packet size	512 bytes
Data transmission rate	1 packet/sec
Initial energy per node	2 Joules
Energy model	Radio energy model (TX = 0.66 J/bit, RX = 0.39 J/bit)
MAC Protocol	IEEE 802.11
Routing protocols compared	RL-SHR (proposed), AODV, DSR
Node failure model	Random node failures (10%–50% of nodes)
Link failure model	Random drop links based on time intervals
Fault detection trigger	ACK timeout, HELLO beacon loss, residual energy threshold
Reinforcement learning algorithm	Q-Learning (ε-greedy policy)
Learning rate (α)	0.1
Discount factor (γ)	0.9
Exploration rate (ε)	Starts at 0.3, decays over time

A random deployment of sensor nodes was simulated in a square sensing area. The sink node was placed either centrally or at one corner to test varied routing paths. Each simulation was run multiple times with different random seeds to ensure statistical reliability. The above table 2 shows the Simulation parameters for this scenario.

Evaluation Metrics-- The following metrics were used to evaluate protocol performance:

- Packet Delivery Ratio (PDR): Ratio of successfully delivered packets to those generated.
- Average end-to-end delay- This is the time taken for packets to reach the sink.
- Energy Consumption- Total energy consumed across the network.
- Network Lifetime- Time until the first node and 50% of nodes die
- Convergence time- Time taken by RL agents to stabilize their Q-values.

Fault injection strategy- To assess the fault tolerance of RL-SHR, controlled node and link failures were introduced during NS2 simulations. Node failure for 10% to 50 % were simulated by randomly deactivating nodes to mimic energy depletion or physical damage. Link failures were modeled by disabling communication between random node pairs at different intervals. These faults triggered the protocols self healing mechanisms using Ack timeouts, hello message loss and energy thresholds. This strategy allowed realistic testing the proposed protocol's adaptability under dynamic and failure prone network conditions.

6. Results and Analysis

We have conducted comprehensive simulations on NS2 simulator under various conditions of node failure and link failure to test the performance of RL-SHR protocol. The results obtained were compared with AODV and DSR protocols using different metrics like PDR, end-to-end delay, energy consumption, network lifetime and Q-value convergence. The proposed protocol outperforms in all the aspects the reliability, energy efficiency and adaptability.

Packet Delivery Ratio (PDR) - The following figure 2 shows the simulation results, depicts that the proposed RL-SHR protocol consistently achieves a higher PDR compared to AODV and DSR, especially under increasing node failure rates.

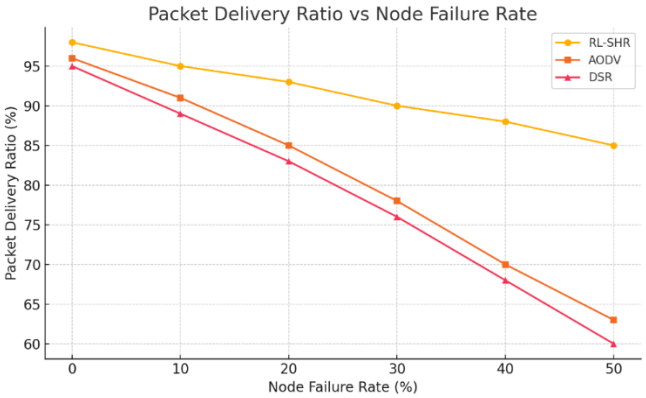


Fig.2. PDR vs node failure rate

At lower failure rates (10% - 20%), all the protocols performs coparably good. However, as the failure rate increases to 30%, 40% and 50%, the performance of RL-SHR maintains a significantly more stable packet delivery ratio. This is due to RL_SHRs ability to learn dynamically and adapt to mode and link failures using Q-learning, which enables it to reroute data efficiently through reliable paths. In contrast, AODV and DSR, which rely on static route discovery mechanisms, struggle to adapt to dynamic changes, leading to a sharp drop in PDR under fault conditions. Overall, the proposed RL-SHR enhances the reliability of message delivery by as much as 20% in high failure cases, thus proving its efficiency as a self-healing and fault-tolerant routing mechanism.

Average End-to-End Delay- End-to-end delay was analyzed through two perspectives: variation with node failure rate and variation over simulation time.

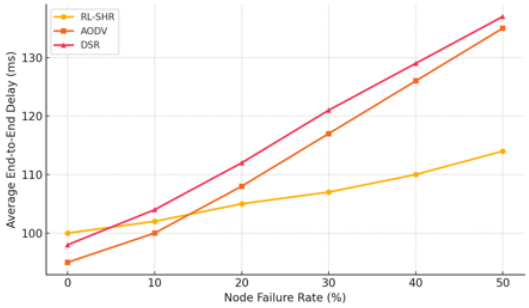


Fig.3a. End to end delay wrt Node failure

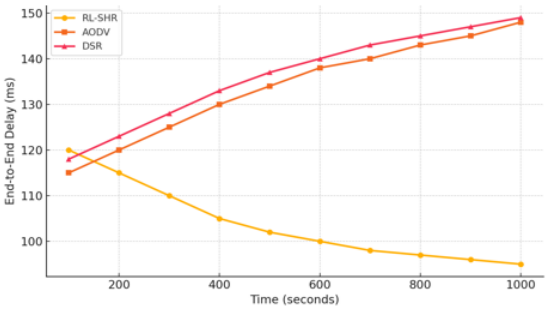


Fig.3b. End to end delay wrt time

As illustrated in Fig 3 (a), RL-SHR achieves lower delay values for all failure rates compared to AODV and DSR. This is because RL-SHR has the capability to adaptively re-route data through trusted nodes based on Q-values, whereas AODV and DSR have the limitation of reactive path discovery, leading to higher delay values for frequent failures.

Fig 3 (b) illustrates that RL-SHR adapts to failures in WSNs. At the beginning, the delay is slightly higher due to exploration in the learning process, but it quickly converges as optimal routes are discovered. On the other hand, AODV and DSR exhibit a sudden increase in delay values after failure occurrences. These observations demonstrate that RL-SHR not only offers fault tolerance but also fast recovery, making it more applicable to dynamic WSNs.

Energy Consumption- Energy efficiency is an important consideration in the development of Wireless Sensor Networks, considering the limited power capabilities of WSN nodes. The proposed protocol performs better in energy management compared to AODV and DSR

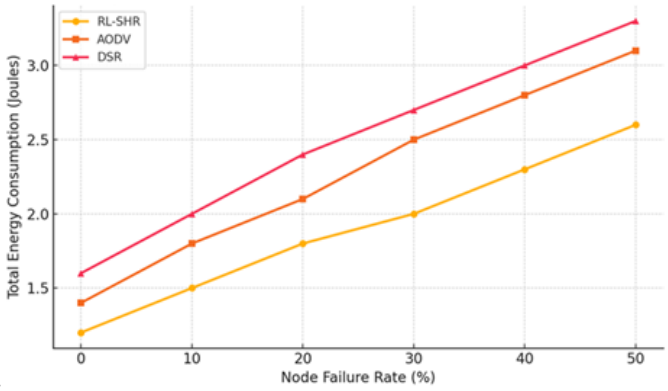


Fig.4. Total energy consumption vs node failure rate

In Fig 4 RL-SHR consistently consumed less energy across all levels of node failure. This is because RL-SHR uses intelligent routing, which skips nodes that are unreliable or have depleted energy, thus eliminating the need for frequent retransmissions and route discoveries. AODV and DSR, on the other hand, used more energy, especially when the failure rate was high due to frequent flooding and route discoveries. Additionally, RL-SHR load-balances the network by routing traffic according to the Q-value, thus preventing nodes from exhausting their capacity prematurely. These observations clearly indicate that RL-SHR not only enhances network fault tolerance but also maximizes the lifetime of the network by avoiding unnecessary energy consumption.

Network Lifetime - Network lifetime was measured using two important parameters: the time taken for the first node to die and the time taken for 50% of the nodes to become inactive. Both parameters indicate the network's energy distribution performance in the early stages and its long-term viability. The proposed RL-SHR protocol always performs

better than AODV and DSR

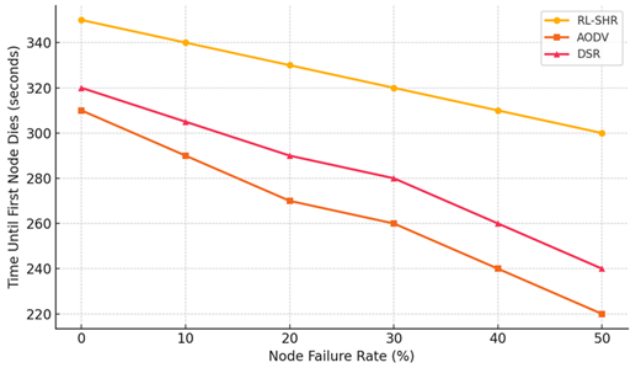


Fig.5. Network lifetime vs node failure rate

The fig 5, shows that RL-SHR has a higher first node death time because of its energy-aware routing strategy, which equally loads traffic on the network, thus avoiding the early death of vital nodes. Additionally, the protocol also has a higher 50% node survival time because of its dynamic adaptation to the network conditions and avoidance of routes that are frequently used and have low energies. On the contrary, AODV and DSR have an imbalance in energy consumption and frequent retransmissions, because of which the nodes die earlier and network performance degrades. The above results show the effectiveness of RLSHR in maintaining network functionality for a longer period despite the rising failure rates.

Learning Convergence- Fig 6 shows the learning convergence of the average Q-values with simulation time in the RL-SHR protocol. At the beginning, the Q-values grow rapidly as the nodes learn different routing behaviors and get feedback from the network. This learning process is characterized by rapid updates based on changing rewards. However, as time passes by, the Q-values tend to become stable, thereby indicating that the nodes have learned the optimal or near-optimal routing strategies. At around 800-1000 simulation seconds, the Q-values tend to converge, thereby indicating that the protocol has adapted to the network conditions and reduced further exploration. The above result clearly indicates that the RL-SHR protocol has learned effective and energy-efficient paths and avoided the possibility of making random routing choices.

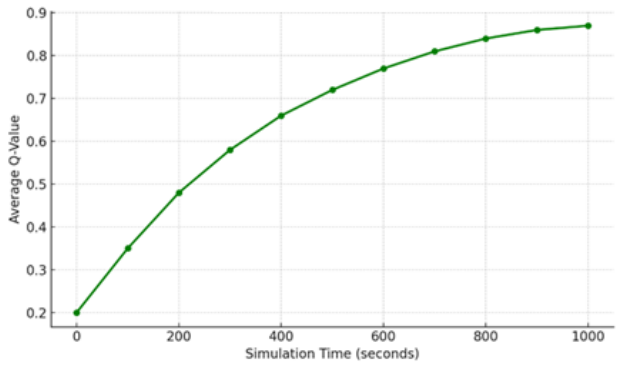


Fig.6. Q value convergence over time

The RL-SHR protocol showed better fault tolerance, adaptability, and energy awareness in the fault-prone WSN environment than the classical routing protocols. The addition of Q-learning helped achieve real-time decision-making and self-healing capabilities, which improved the network lifetime and reliability in adverse environments. These findings confirm the effectiveness of the RL-SHR protocol as a reliable solution for future WSN applications in dynamic and critical environments.

7. Conclusions and Future Work

This paper introduces the RL-SHR, a Reinforcement Learning based Self-Healing Routing protocol to enhance the fault tolerance and energy efficiency of Wireless Sensor Networks. The RL-SHR protocol employs the Q-Learning algorithm at each node to facilitate a decentralized, adaptive routing mechanism according to real-time observations of link quality, residual energy, and delivery success. The proposed protocol includes a reward-driven learning model and a self-healing module to ensure reliable communication in the event of node and link failures. Simulation experiments have clearly shown that the RL-SHR protocol outperforms traditional routing protocols such as AODV and DSR in several performance aspects. In particular, the RL-SHR protocol maintained a packet delivery ratio of above 85% even at a 50%

node failure rate, while AODV and DSR decreased to below 70% and 60%, respectively. Furthermore, the RL-SHR protocol reduced the end-to-end delay and energy consumption by adaptively avoiding congested and low-energy nodes. In the aspect of network lifetime, the RL-SHr protocol improved the time to first node death and 50% nodes death compared to AODV by more than 25% in different scenarios. Moreover, the convergence analysis indicates that the RL-SHR protocol converges very quickly to its Q-values, ensuring efficient routing operations over time, which is not feasible in traditional reactive routing protocols.

These comparative observations reaffirm that RL-SHR provides a significant performance benefit over non-learning-based solutions, particularly within dynamic and failure-prone WSN settings. Although AODV and DSR routing solutions are based on static path discovery and maintenance, the RL-SHR approach dynamically learns and adapts, thus being more robust, energy-efficient, and scalable. This research effort makes a contribution towards making WSNs more robust, autonomous, and amenable for mission-critical applications. Although the performance benefits in terms of packet delivery, delay, energy efficiency, and network lifetime have been shown to be significant, there exist some limitations. The existing performance assessment has been carried out on networks of moderate size; scalability to thousands of nodes has not been investigated yet, and it may cause convergence-related delays and increased computational complexity. Furthermore, the existing investigation has considered a mostly static network topology, and mobility-related WSN settings, such as vehicular or UAV-based WSNs, have not been investigated adequately.Addressing these aspects will be crucial for extending RL-SHR to large-scale and mobile deployments. Future work will therefore investigate scalability optimizations, mobility-aware learning strategies, and real-world validation on hardware testbeds. The future work will also explore extensions such as mobility aware RL models, the integration of deep reinforcement learning e.g., DQN, and cooperative multi agent learning to further enhance convergence & scalability.

Author Contributions Statement

Dipti Chauhan contributed to the conceptualization of the research, design of the RL-SHR protocol, supervision of the study, overall framework development and implementation on the NS2 simulation environment.
Pritika Bahad, contributed to the mathematical formulation of the Q-learning model, conducted performance evaluation, performed data analysis, and contributed to manuscript drafting.
Jay Kumar Jain validation of results, critical review of the manuscript, and refinement of the theoretical framework.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare that they have no conflicts of interests.

Funding Declaration

This work does not receive external funding.

Data Availability Statement

The data supporting the findings of this study were generated using NS2 simulations. Simulation scripts and generated datasets are available from the corresponding author upon reasonable request.

Ethical Declarations

This work does not involve human participants, animals, or any biological subjects. Therefore, ethical approval and informed consent were not required.

Acknowledgements

We sincerely thank the experts for their professional evaluation and valuable recommendations, which have contributed to improving the quality of the experiment and the reliability of its results.

Declaration of Generative AI in Scholarly Writing

In the preparation of this manuscript, we have used used AI-assisted language tools to improve grammar, clarity, and overall presentation. The research design, algorithm development, NS2 simulation implementation, experimental evaluation, and interpretation of results were conducted entirely by the authors. The authors take full responsibility for the scientific content and integrity of this work.

Abbreviations

The following abbreviations are used in this manuscript:

AI – Artificial Intelligence
AODV – Ad hoc On-Demand Distance Vector
CBR – Constant Bit Rate

DL – Deep Learning
DQN – Deep Q-Network
DSR – Dynamic Source Routing
IoT – Internet of Things
MAC – Medium Access Control
NS2 – Network Simulator 2
PDR – Packet Delivery Ratio
QoS – Quality of Service
RL – Reinforcement Learning
RL-SHR – Reinforcement Learning-Based Self-Healing Routing
UAV – Unmanned Aerial Vehicle
WSN – Wireless Sensor Network

References

[1] D. Chauhan, J. Jain, and P. Bahad, "Performance Evaluation of 802.11a/G Wireless Networks With Ip6hc," Journal of Management Information and Decision Sciences, vol. 24, no. S3, pp. 1–7, 2021, Accessed: Mar. 05, 2026. [Online]. Available: https://www.abacademies.org/articles/performance-evaluation-of-80211ag-wireless-networks-with-ip6hc.pdf

[2] Y. Huang, J.-F. Martínez, J. Sendra, and L. López, "ResilientWireless Sensor Networks Using Topology Control: A Review," Sensors, vol. 15, no. 10, pp. 24735–24770, Sep. 2015, doi: https://doi.org/10.3390/s151024735.

[3] J. K. Jain and D. Chauhan, "Optimized secure and energy-efficient approach for IoT-enabled wireless sensor networks," Pervasive and Mobile Computing, vol. 110, p. 102049, May 2025, doi: https://doi.org/10.1016/j.pmcj.2025.102049.

[4] D. Chauhan and S. Sharma, "Performance Evaluation of Different Routing Protocols in IPv4 and IPv6 Networks on the Basis of Packet Sizes," Procedia Computer Science, vol. 46, pp. 1072–1078, 2015, doi: https://doi.org/10.1016/j.procs.2015.01.019.

[5] H. Fang et al., "Self-Healing in Knowledge-Driven Autonomous Networks: Context, Challenges, and Future Directions," IEEE Network, pp. 1–1, 2024, doi: https://doi.org/10.1109/mnet.2024.3416850.

[6] K. Sivamayil, E. Rajasekar, B. Aljafari, S. Nikolovski, S. Vairavasundaram, and I. Vairavasundaram, "A Systematic Study on Reinforcement Learning Based Applications," Energies, vol. 16, no. 3, p. 1512, Jan. 2023, doi: https://doi.org/10.3390/en16031512.

[7] M. Ganjalizadeh, H. S. Ghadikolaei, A. Azari, A. Alabbasi, and M. Petrova, "Saving Energy and Spectrum in Enabling URLLC Services: A Scalable RL Solution," IEEE Transactions on Industrial Informatics, vol. 19, no. 10, pp. 10265–10276, Oct. 2023, doi: https://doi.org/10.1109/tii.2023.3240592.

[8] V. Rajagopal, B. Velusamy, and S. Rathinasamy, "Double Q-learning-based adaptive trajectory selection for energy-efficient data collection in wireless sensor networks," International Journal of Communication Systems, vol. 36, no. 7, Feb. 2023, doi: https://doi.org/10.1002/dac.5452.

[9] A. Keerthika and V. Berlin Hency, "Reinforcement-Learning based energy efficient optimized routing protocol for WSN," Peer-to-Peer Networking and Applications, Mar. 2022, doi: https://doi.org/10.1007/s12083-022-01315-6.

[10] T. Zhang, K. Zhu, and E. Hossain, "Data-Driven Machine Learning Techniques for Self-Healing in Cellular Wireless Networks: Challenges and Solutions," Intelligent Computing, vol. 2022, Jan. 2022, doi: https://doi.org/10.34133/2022/9758169.

[11] D. Godfrey, B. Suh, B. H. Lim, K.-C. Lee, and K.-I. Kim, "An Energy-Efficient Routing Protocol with Reinforcement Learning in Software-Defined Wireless Sensor Networks," Sensors, vol. 23, no. 20, p. 8435, Jan. 2023, doi: https://doi.org/10.3390/s23208435.

[12] N. Alsalmi, K. Navaie, and H. Rahmani, "Energy and throughput efficient mobile wireless sensor networks: A deep reinforcement learning approach," IET networks, May 2024, doi: https://doi.org/10.1049/ntw2.12126.

Authors' Profiles



Dr. Dipti Chauhan: currently serves as the Professor and Head of the Department of Artificial Intelligence & Data Science at PIEMR, Indore. She earned her Ph.D. from Maulana Azad National Institute of Technology (MANIT), Bhopal, specializing in Next Generation Networks and IPv6. During her doctoral studies, she was awarded a prestigious fellowship by the Ministry of Human Resource Development (MHRD), Government of India. Dr. Chauhan is an IPv6 Certified Gold and Silver Network Engineer, accredited by the IPv6 Forum at University Sains Malaysia. Her core research interests span a wide range of advanced domains, including Data Mining and Warehousing, Artificial Intelligence, Machine Learning, Data Science, Next Generation Networks, and the Internet of Things (IoT).



Dr. Pritika Bahad: is a seasoned academic with over 14 years of teaching experience. She completed her M.Tech in Mobile Computing Technology and PhD in Computer Science from Devi Ahilya University, Indore. Currently serving as a Senior Assistant Professor in the Department of Artificial Intelligence and Data Science, Dr. Bahad has made significant contributions through 20+ book chapters and research papers in international and national journals and conferences. Additionally, Dr. Bahad has actively contributed to various hackathons and e-learning initiatives.



Dr. Jay Kumar Jain is currently working as an Assistant Professor in the Department of Mathematics, Bioinformatics and Computer Applications, MANIT, Bhopal. He did his Ph. D. from Maulana Azad National Institute of Technology, Bhopal in 2015. He was awarded with Research Fellowship by MHRD for completing his Ph.D. He has research as well as teaching experience of about 13 years. He has published around 55+ research papers and book chapters in SCI/SCOPUS/Referred International/National Journal and Conferences He has also authored 03 books for International Publisher. He has also been granted 3 International patents and 1 Indian Patent. 3 Indian design patents and 2 copyrights in registered IPR, India. He has been a reviewer in many international journals/conferences including Elsevier, IEEE Access, and Springer. He has lifetime membership of various professional societies such as IEEE, CSI, Franklin, IDES, IAENG, SDIWC, and many more. His research interests include Wireless Sensor Networks, the Internet of Things, and Mobile Ad hoc Networks.

How to cite this paper: Dipti Chauhan, Pritika Bahad, Jay Kumar Jain, "Reinforcement Learning-Based Self-Healing Routing in Fault-prone Wireless Sensor Networks", International Journal of Computer Network and Information Security(IJCNIS), Vol.18, No.2, pp.112-125, 2026. DOI:10.5815/ijcnis.2026.02.07