electronics

MDPI

Article

Research on High-Efficiency Routing Protocols for HWSNs Based on Deep Reinforcement Learning

Yu Song [1,2,*], Zhigui Liu [1,*], Kunran Li [3], Xiaoli He [4] and Weizhuo Zhu [4]

[1] School of Information Engineering, South West University of Science and Technology, Mianyang 621010, China
[2] Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things, Sichuan University of Science and Engineering, Zigong 643000, China
[3] School of Cybersecurity, Chengdu University of Information Technology, Chengdu 610000, China; likunran@suse.edu.cn
[4] School of Computer Sciencen, Sichuan University of Science and Engineering, Zigong 643000, China; hexiaoli_suse@hotmail.com (X.H.); 323085406137@stu.suse.edu.cn (W.Z.)
* Correspondence: songyu@suse.edu.cn (Y.S.); hexiaoli@suse.edu.cn (Z.L.); Tel.: +86-159-8416-9508 (Y.S.)

Abstract: In heterogeneous wireless sensor networks (HWSNs), optimizing energy efficiency presents significant challenges due to variations in node energy levels and the complexity of the network environment. This paper introduces an energy efficiency optimization algorithm for HWSNs based on the Deep Q-Network (HDQN). The algorithm aims to address these challenges by selecting the optimal information transmission path. The HDQN leverages energy differences between nodes and real-time environmental data to enhance network efficiency. Its reward function takes into account node distance, remaining energy, and relay node count to balance node participation and minimize overall energy consumption. The Deep Q-Network (DQN) uses the mean squared error for precise reward estimation, and an improved packet header structure supports effective routing decisions. Simulation results show that the HDQN significantly outperforms existing algorithms—EEHCHR, 2L-HMGEAR, NCOGA, DEEC, and SEP—in terms of energy efficiency, network lifetime, and robustness, demonstrating its potential to advance the performance of HWSNs. The research results of the paper provide a theoretical basis for future energy efficiency research in wireless communication and contribute to the study of the new generation of wireless networks.

Keywords: heterogeneous wireless sensor networks; Deep Q-Network (DQN); energy efficiency; path optimization; reward function; routing algorithm

Citation: Song, Y.; Liu, Z.; Li, K.; He, X.; Zhu, W. Research on High-Efficiency Routing Protocols for HWSNs Based on Deep Reinforcement Learning. *Electronics* **2024**, *13*, 4746. https://doi.org/10.3390/electronics13234746

Academic Editor: Christos J. Bouras

Received: 24 September 2024
Revised: 20 November 2024
Accepted: 26 November 2024
Published: 30 November 2024

1. Introduction

Wireless sensor networks (WSNs) are a new type of network that integrates the various technologies of sensors, computers, communications, and other multidisciplinary information acquisition and processing. With the rapid development of Internet of Things (IoT) technology, WSNs have been widely used in environmental monitoring, smart homes, medical health and other fields, and have a wide range of application prospects in industrial and military fields. According to whether the nodes in the network have the same function, WSNs can be divided into homogeneous WSNs and heterogeneous WSNs (HWSNs). The nodes of homogeneous WSNs have the same type and function, making them suitable for single data collection and processing. HWSNs can be composed of a variety of different types and functions of sensors, such as temperature and humidity sensors, combustible gas sensors, infrared sensor modules, and flame sensor modules to adapt to diverse data needs and environmental changes. Current research on WSNs mainly focuses on the field of homogeneous WSNs. Unlike homogeneous WSNs, sensor nodes in HWSNs have different resource configurations, such as computing power, communication capacity, storage capacity, and energy supply. This diversity can better meet the needs of practical

application scenarios. At the same time, due to the limited energy sources for sensor nodes, extending their lifecycle while ensuring reliable transmission has become a key issue. Therefore, designing efficient routing protocols is an important topic for studying energy efficiency.

1.1. Related Work

Research on HWSNs has yielded significant results, focusing primarily on energy efficiency optimization, load balancing, Quality of Service (QoS) assurance, routing protocol design, and energy-efficient communication strategies.

The progress of HWSNs has been profoundly impacted by the demand for energy efficiency, especially in large-scale deployments where sensor nodes are mainly battery-powered. The heterogeneity of these networks, in terms of both node capabilities and mobility, presents distinctive challenges that require innovative solutions to optimize energy consumption and guarantee network robustness. This section examines the recent literature that tackles these challenges through novel routing protocols, deployment strategies, and energy management techniques. Regarding energy-efficient routing protocols, Li et al. introduced the NMSFRA routing protocol, designed for HWSNs characterized by mobile nodes. Distinct from traditional homogeneous WSN routing protocols, NMSFRA aimed to balance energy consumption by addressing the uneven cluster distribution and unstable network connections resulting from node mobility. This protocol effectively accommodated the dynamic nature of heterogeneous nodes, thereby enhancing overall energy efficiency and network stability [1]. Another significant contribution in this field was the energy-efficient cooperative routing scheme (EERH) proposed in reference [2]. The EERH enhanced data transmission efficiency in HWSNs by enabling different WSNs to share routing paths and nodes for event message forwarding. The routing strategy in the EERH was dynamically adjusted based on the remaining energy of the underlying sensors and their neighboring sensors, reducing transmission energy consumption and extending the network lifespan [2].

In terms of deployment optimization, reference [3] investigated the deployment of heterogeneous fusion centers (FCs) in WSNs, modeling the optimal placement of access points (APs) and FCs as an optimization problem with the objective of minimizing total wireless communication power consumption. The study considered both static and mobile WSN scenarios, emphasizing the need for adaptive deployment strategies in heterogeneous environments. In terms of clustering and energy management, the significance of efficient clustering techniques in prolonging the lifespan of HWSNs was emphasized in references [4,5]. Reference [4] proposed a distributed energy-based epoch-clustering method combined with a multi-parameter weighted scalarization function to optimize cluster head selection. This approach introduced a novel weight calculation strategy using the analytical hierarchy process (AHP) and a two-phase analytical algorithm, enhancing cluster head selection efficiency. Similarly, reference [5] presented the cluster routing protocol for heterogeneous network (CPHN), which selected cluster heads based on both initial and remaining energy levels to maximize energy efficiency. This protocol ensured prolonged network operation by prioritizing nodes with higher energy reserves.

In terms of performance analysis and enhancement, reference [6] critically assessed various fixed heterogeneous clustering algorithms, evaluating their performance in terms of network lifespan and throughput in mobile node environments. The proposed cluster head-restricted energy-efficient protocol for WSNs (CREW) modified channel selection thresholds in two-layer HWSNs to improve network survival time, overcoming the limitations observed in traditional fixed clustering algorithms. In addition, reference [7] addressed specific constraints in heterogeneous 5G WSNs, focusing on reducing costs and energy consumption and ensuring reliable data transmission. The proposed routing strategy aimed to alleviate challenges such as excessive energy consumption in void areas, packet loss, and the over-consumption of energy, enhancing network stability.

Finally, reference [8] reconsidered the concept of the EERH, highlighting its application in forming a heterogeneous sensor network where multiple WSNs share resources for data forwarding. This approach dynamically established routing paths based on the direction of event data packet transmission and the energy levels of underlying sensors and their neighbors, further aggregating data packets for the same direction to conserve energy. The reviewed literature collectively emphasizes the crucial role of energy efficiency in HWSNs, offering a range of solutions from advanced routing protocols and deployment strategies to innovative clustering and energy management techniques. These contributions not only address immediate challenges such as node mobility and energy consumption but also lay the foundation for future research in optimizing the performance and longevity of HWSNs.

Overall, these contributions underscore the ongoing advancements in HWSNs research, highlighting the integration of optimization techniques, routing strategies, and security measures to address the complex challenges faced by modern sensor networks. However, despite significant advancements in HWSNs research, which highlight the integration of optimization techniques, routing strategies, and security measures, several challenges persist. Issues such as the validation of real-world effectiveness, increased system complexity, the trade-off between energy consumption and performance, and the need for enhanced security and privacy protection remain. Additionally, the adaptability of these systems in dynamic environments and the potential for effective cross-layer comprehensive optimization require further investigation. Addressing these challenges is crucial for achieving more efficient and reliable sensor networks.

1.2. Contribution

In this study, an energy-efficient routing strategy for HWSNs based on the Deep Q-Network (DQN) is proposed. The strategy is validated for its effectiveness in optimizing energy efficiency and extending the network lifetime through system modeling, algorithm design, simulation experiments, and performance analysis. The specific contributions are as follows:

Firstly, an intelligent routing strategy based on the DQN is introduced to dynamically optimize energy efficiency in HWSNs.

Secondly, a MATLAB simulation environment is established to verify the performance of the algorithm under different application scenarios.

Finally, through detailed experiments and simulation analyses, the advantages of the DQN-based routing strategy in reducing energy consumption, extending network lifetime, and improving data transmission performance are demonstrated.

1.3. Organization

The remainder of this paper is organized as follows. Section 2 presents the application scenario and system model. Section 3 explores optimization problems, including Q-learning and DQN. Section 4 introduces an improved DQN algorithm. Numerical results are presented in Section 5. Finally, Section 6 concludes this paper.

2. Application Scenario and System Model

2.1. Application Scenario Analysis

The application of HWSNs in smart agriculture significantly enhances the efficiency of soil nutrient monitoring and management, environmental monitoring, weather forecasting, and crop growth monitoring. As illustrated in Figure 1, HWSNs deployed in agricultural fields utilize a variety of sensor nodes to continuously gather data on soil moisture, PH values, nutrient content, light exposure, and air temperature. Cluster heads (CHs) aggregate this monitoring data and employ an optimized algorithm to select the most efficient relay nodes for data transmission to the base station (BS). The collected data are subsequently uploaded to a smart cloud platform, enabling farmers to access and analyze the information via remote computers or mobile devices. This capability allows farmers to adjust irrigation and fertilization strategies in real time. Through the

real-time data monitoring and precise management facilitated by HWSNs, farmers can more scientifically manage the agricultural environment, thereby improving crop yield and quality. This not only propels the development of smart agriculture but also provides technological support for agricultural modernization. However, HWSNs require regular maintenance and battery replacement, which increases maintenance costs. To enhance the energy efficiency of HWSNs, it is essential to implement optimization strategies, such as employing information transmission energy efficiency optimization algorithms to extend the network's operational lifetime. This forms the primary focus of the research presented in this paper.

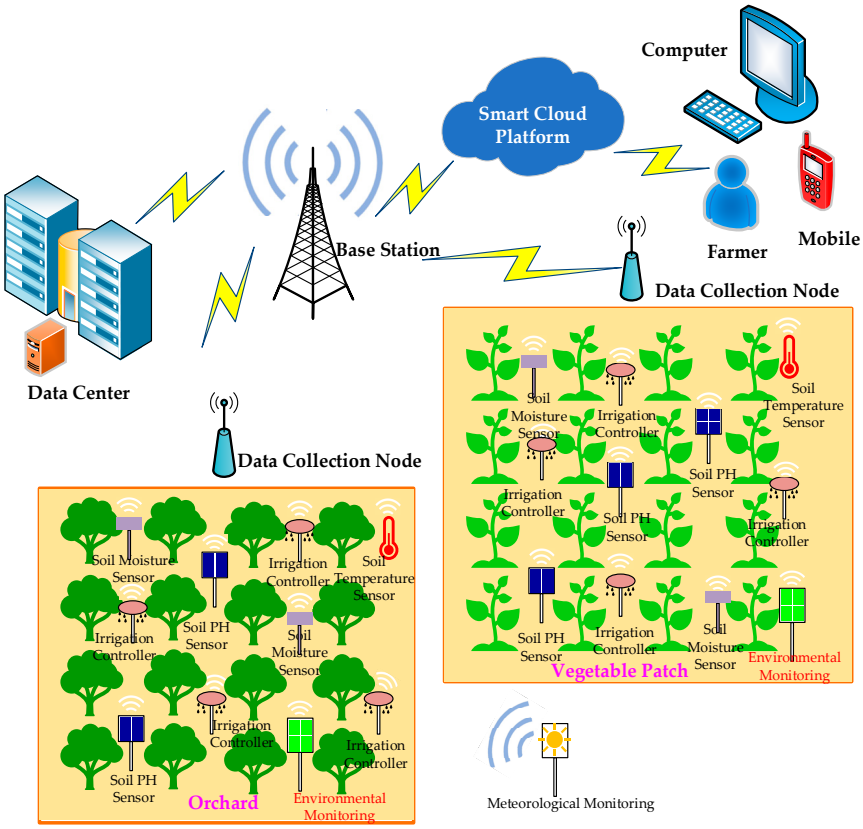

Figure 1. Application scenario of HWSNs in smart agriculture.

In the smart agriculture application scenario depicted in Figure 1, various types of sensor nodes need to be deployed across extensive agricultural areas and their continuous and stable operation ensured over extended periods. In practical applications, problems such as uneven energy consumption among heterogeneous nodes and premature energy depletion, leading to monitoring blind spots, may emerge. This necessitates the consideration of energy efficiency coordination management strategies for HWSNs. Given its lightweight and easy-to-implement characteristics, the DQN is particularly suitable for addressing energy efficiency coordination management issues in networks with relatively scarce hardware resources, such as HWSNs.

2.2. Network Topology

The system model of HWSNs is illustrated in Figure 2. In Figure 2, it is assumed that sensor nodes are randomly distributed within the area. The BS has unlimited energy and is located at the center of the monitoring area. Among the N heterogeneous sensor nodes, there are three types of nodes: super nodes, advanced nodes, and ordinary nodes. In HWSNs, the ordinary nodes include various sensors such as soil temperature and moisture sensors, soil PH sensors, and irrigation controllers. These nodes are responsible for data

collection. Due to the different functionalities and energy consumption characteristics of these sensors, the network topology modeling becomes complex. As shown in Figure 2, to simplify the topology of HWSNs, clustering algorithms are typically employed to group ordinary nodes into different clusters.

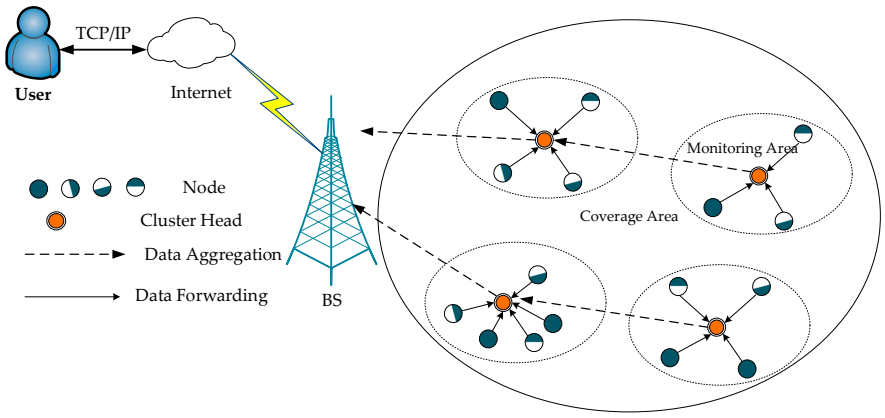

Figure 2. System model of HWSNs.

In addition to ordinary nodes, HWSNs also include a certain number of super nodes. These super nodes possess advanced data processing capabilities and ample initial energy resources, which effectively support the stable operation and efficient data processing of the network. Super nodes generally serve as cluster heads or relay nodes, handling data collection, fusion, and communication tasks. Ordinary sensor nodes primarily focus on data collection and transmit the collected data to the cluster heads. The cluster head nodes aggregate and process the received data and subsequently transmit them to the users via the BS. This hierarchical design enhances network efficiency and data transmission by utilizing super nodes for data processing and network management. The robust data processing capabilities and sufficient energy resources of super nodes enable them to play a crucial role in scenarios with large data volumes and complex conditions. By judiciously distributing tasks among super nodes, the network load can be balanced and resources optimally utilized, further improving the overall network performance.

2.3. Node Model and Energy Model

Among the N nodes, there are N_{su} super nodes, N_{ad} advanced nodes, and N_{or} ordinary nodes. The advanced nodes have several times more energy than the ordinary nodes. The proportions of super nodes and advanced nodes are M_s and M_a, respectively. The number of each type of node can be expressed by the following equation:

$$N_{su} = N \times M_s \tag{1}$$

$$N_{ad} = N \times M_a \tag{2}$$

$$N_{or} = N \times (1 - M_s - M_a) \tag{3}$$

where N_{su}, N_{ad}, and N_{or} represent the number of super nodes, advanced nodes, and ordinary nodes, respectively.

Assume that the energy of a super node is β_{su} times that of an ordinary node, and the energy of an advanced node is β_{ad} times that of an ordinary node. The energy consumption of an ordinary node includes energy expenditures for data acquisition, data processing, and data transmission. The total energy consumption of the i-th node can be expressed by the following equation:

$$E^i_{total} = E^i_{sense} + E^i_{proc} + E^i_{trans} \tag{4}$$

where E^i_{sense} represents the energy consumption for data sensing, E^i_{proc} denotes the energy consumption for data processing and E^i_{trans} indicates the energy consumption for data transmission.

The total initial energy for ordinary nodes can be expressed as

$$E_{total\_or} = N_{or} \times E \tag{5}$$

where E is the initial energy of a single ordinary node. The initial energy of super nodes is β_{su} times that of ordinary nodes. Consequently, the total initial energy of super nodes can be expressed as

$$E_{total\_su} = N_{su} \times \beta_{su} \times E \tag{6}$$

where $\beta_{su} \times E$ is the initial energy of a single super node.

The initial energy of advanced nodes is β_{ad} times that of ordinary nodes. Therefore, the total initial energy of advanced nodes can be represented as

$$E_{total\_ad} = N_{ad} \times \beta_{ad} \times E \tag{7}$$

where $\beta_{su} \times E$ is the initial energy of a single advanced node.

The total initial energy of the HWSNs can be calculated as

$$E_{total} = E_{total\_or} + E_{total\_su} + E_{total\_ad} \tag{8}$$

3. Optimization Problem Analysis

With the node model and energy model defined, we can proceed to formulate and analyze the optimization problem for data transmission within the HWSNs. The objective is to optimize the network's performance by selecting the best routes for data transmission, thereby extending the overall network lifetime and minimizing energy consumption.

3.1. Problem Formulation

The optimization problem can be defined as a single-objective optimization problem with multiple constraints. The objective function aims to minimize energy consumption. This can be achieved by reducing the total energy used by the nodes during data transmission and extending the network's operational time through balanced energy consumption across all nodes. This optimization objective can be mathematically expressed as follows:

$$\min E^{ec}_{total} = \sum_{t=1}^{T} \sum_{i=1}^{N_{or}} E^i_{total\_or}(t) + \sum_{t=1}^{T} \sum_{j=1}^{N_{su}} E^j_{total\_su}(t) + \sum_{t=1}^{T} \sum_{k=1}^{N_{ad}} E^k_{total\_ad}(t) \tag{9}$$

where $E^i_{total\_or}(t)$, $E^j_{total\_su}(t)$, and $E^k_{total\_ad}(t)$ respectively represent the total energy consumption by the i-th ordinary node, j-th super node, and k-th advanced node at time t.

3.2. Constraints

The optimization is subject to several constraints, e.g., energy constraints. Each node has limited initial energy, and the total energy consumption should not exceed this initial energy.

$$\sum_{t=1}^{T} \sum_{i=1}^{N_{or}} E^i_{total\_or}(t) \leq E_{total\_or} \tag{10}$$

$$\sum_{t=1}^{T} \sum_{j=1}^{N_{su}} E^j_{total\_su}(t) \leq E_{total\_su} \tag{11}$$

$$\sum_{t=1}^{T} \sum_{k=1}^{N_{ad}} E^k_{total\_ad}(t) \leq E_{total\_ad} \tag{12}$$

3.3. Algorithm Design

3.3.1. Overview of Q-Learning

Reinforcement learning (RL) is one of the mainstream intelligent methods used to address data transmission and routing optimization issues in WSNs. Reinforcement learning involves an Agent interacting with the environment to take actions and obtain the maximum cumulative reward, thereby continuously optimizing its decision-making capability. Reinforcement learning methods are defined by a quadruple (S, A, P, r), where S represents the current state of the Agent; A denotes the action taken by the Agent in the current state; P indicates the probability of transitioning from the current state to other states after executing an action; and r signifies the reward received by the Agent after performing the corresponding action. Typically, the reinforcement learning process can be described using a Markov decision process (MDP).

Q-learning is a model-free reinforcement learning algorithm used to solve optimization problems by learning the optimal action selection policy for a given environment. It is a type of temporal difference learning aimed at finding the optimal policy by estimating the value of actions in various states of the environment. The Q-value update algorithm is as follows:

$$Q(S, A) \leftarrow Q(S, A) + \alpha \Big[R + \gamma \max_{a} Q(S', a) - Q(S, A) \Big] \tag{13}$$

In Equation (13), α represents the learning rate, γ denotes the reward decay factor, and R is the reward for action A. Equation (13) signifies that the Agent uses the past Q-values from the Q-table as experience to estimate the updated Q-values for each possible action in the next state S'. It then selects the maximum Q-value, multiplied by the decay factor γ, and adds the actual reward value to compute the current Q-value update. When γ is close to 0, the Agent relies predominantly on existing experience and does not learn new information, which is known as a greedy strategy. Conversely, when γ is close to 1, the Agent relies more on new information rather than prior experience, leading to a more exploratory approach, though this can potentially cause instability in the learning process. When γ is close to 0, only the immediate reward is considered, whereas a value close to 1 emphasizes long-term high returns. If the Q-values reach or exceed 1, there is a risk of divergence. The detailed initialization and update mechanism of the Q-function (Q-Matrix) can be described in detail as follows.

The Q-function uses the Bellman equation, accepting two inputs, namely the state and the action, described as follows:

$$Q_{\pi}(s, a) = E_{\pi} \big[r' + \gamma r'' + \gamma r''' + \cdots | s, a \big] \tag{14}$$

where $Q_{\pi}(s, a)$ represents the Q-value of the given state and $E_{\pi}[r' + \gamma r'' + \gamma r''' + \cdots | s, a]$ represents the expected discounted cumulative reward of the given state and action. The update method for Q-value based on the principle of the Bellman equation is as follows:

$$Q^*(s, a) = (1 - \delta) Q(s, a) + \delta r + \gamma \max_{a} Q(s', a) \tag{15}$$

where $Q^*(s, a)$ represents the Q-value of the next state and δ represents the learning rate. $Q(s, a)$ represents the current Q-value. $\max_{a} Q(s', a)$ represents the estimation of the future optimal value. By iteratively updating the Q-values, the DQN algorithm approximates the Q-values, using DNNs to learn the optimal action policy, thereby facilitating effective decision-making in reinforcement learning tasks.

3.3.2. Deep Reinforcement Learning

Q-learning is a value iteration-based reinforcement learning algorithm that finds the optimal policy by continually updating the Q-values of state–action pairs. The original Q-learning algorithm is designed to handle finite and discrete action spaces. When faced with a vast number of state and action combinations, finding the action that maximizes

the Q-value becomes exceptionally challenging, as this requires operations that involve evaluating or maximizing over an infinite number of possible actions, which is impractical in real-world scenarios. The process of deep reinforcement learning is illustrated in Figure 3.

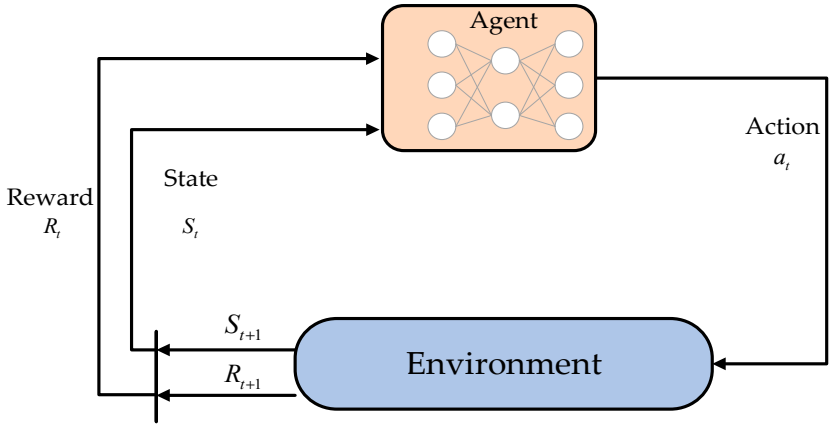

Figure 3. Deep reinforcement learning process.

The target network can improve the training stability of the DQN. Its main function is to select actions based on the maximum value of the current network $Q(S, a)$ during DQN training, and then calculate the value y of the target network based on the maximum Q-value $Q'(Q', a)$ of the next state S'. However, directly using the same current network for calculations can lead to unstable training. If the Q' value changes too quickly, the y value will also increase accordingly. Therefore, an additional target network is used to calculate Q' to ensure the stability of model training. The experience replay mechanism is employed to stabilize the training process. This involves using an experience buffer to store the states and actions the Agent acquires while interacting with the environment. During training, a random batch of samples is drawn from this buffer for learning. By storing all historical states and actions, this mechanism allows the information to be reused multiple times to optimize the DQN, thereby improving data efficiency. When calculating the target Q-value, the target network is used to calculate the maximum Q-value of the next state.

$$y = r + \gamma \max_{a'} Q_{\text{target}}(s', a') \tag{16}$$

where $Q_{\text{target}}(s', a')$ is the Q-value of action a' selected by the target network in the next state s'.

The calculation equation for the loss function of the DQN algorithm is as follows:

$$L(\theta) = \mathbb{E}\left(R + \gamma \max_{a'} Q(s', a'; \theta) - Q(s, a; \theta) \right)^2 \tag{17}$$

where the loss function defined in Equation (17) represents the difference between the target Q-value and the current Q-value. Once the loss function is obtained, the gradient descent method can be directly applied to optimize the weight parameters θ of the convolutional neural network. In the DQN training process, to improve the efficiency and stability of learning, the DQN algorithm randomly samples pairs of states and actions from the historical dataset for training each time the Q-value is updated. Additionally, using two deep neural networks (DNNs) with the same structure but different parameters helps reduce the likelihood of oscillations and divergences during training, thereby enhancing the stability of the algorithm.

3.3.3. Applying DQN in HWSNs

As shown in Figure 4, a super node is designated as the Agent, which is responsible for sensing the environment of the HWSNs and learning the optimal strategy.

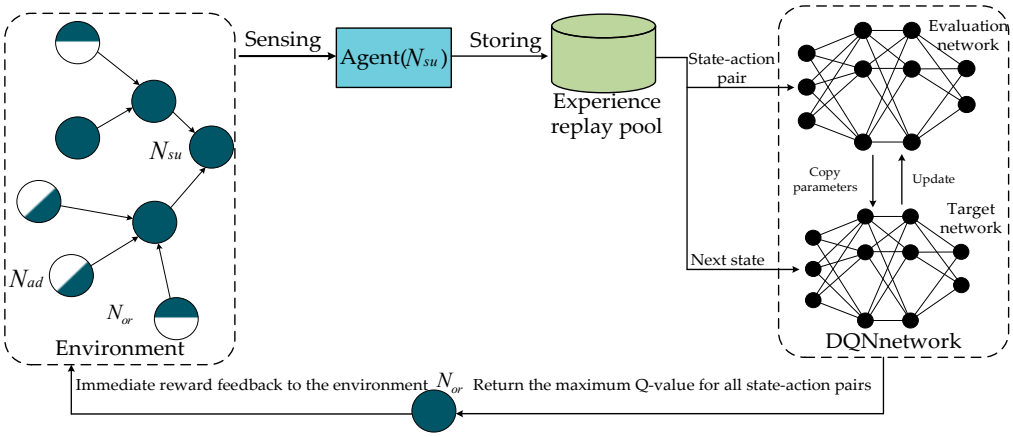

Figure 4. HWSNs system model for clustering data transmission based on DQN.

The Agent stores previous experiences in a replay buffer, which includes information such as states, actions, and rewards. During each training session, a small batch of data is randomly sampled from the stored experiences to train and update the weights of both the evaluation network and the target network in the deep neural network (DNN). The target network transfers the parameters after training to the evaluation network for optimization. During this updating process, the DNN continuously improves the strategy by learning the historical maximum Q-values from the replay buffer, thereby achieving the maximum long-term reward.

The HWSNs are represented by a set $W = (\Phi_N, C, E, Q, H)$, where $\Phi_N = \{s_1, \ldots s_n\}$ represents the set of heterogeneous sensor nodes randomly deployed in the monitoring area of the wireless sensor network. $C = \{c_1, \ldots c_{N_{CH}}\}$ represents the set of cluster heads in the monitoring area. $E = \{e_1, \ldots e_n\}$ represents the set of initial energy levels of each heterogeneous node. $Q = \{q_1, \ldots q_n\}$ represents the set of heterogeneous nodes whose distance to the BS is less than their own transmission threshold. $H = \{h_1, \ldots h_n\}$ represents the set of neighboring nodes of each heterogeneous node.

Assume the following properties for the HWSNs model:

1. Sensor nodes are energy-heterogeneous, but they have identical capabilities in terms of computation, communication, and storage;
2. Each heterogeneous node is equipped with a GPS device;
3. After deployment, the heterogeneous nodes remain static, and each node can belong to only one cluster in each round;
4. Each heterogeneous node can record its own information, including its unique ID, the ID of the cluster head it belongs to in the current round, remaining energy, and the distance to surrounding neighbors;
5. In dynamic scheduling mechanisms, advanced nodes are responsible for assessing network conditions and making corresponding scheduling decisions to ensure the efficient operation of the network.

In HWSNs applications, nodes typically have limited resources, and the network environment is constantly changing. This necessitates that information transmission energy efficiency optimization algorithms for HWSNs possess characteristics such as a high resource utilization efficiency and strong robustness. The DQN algorithm, with the following features, is well suited for addressing the energy efficiency optimization problem in information transmission within HWSNs:

1. Lightweight design: The DQN algorithm employs a small-scale DNN structure, making it suitable for lightweight application environments in HWSNs, where computational, storage, and energy resources are constrained;
2. Adaptability: The reinforcement learning framework within the DQN algorithm offers a self-improvement mechanism, enabling it to learn optimal strategies through interaction with the environment. This is particularly important for handling the dynamic interactions between nodes in a heterogeneous wireless sensor network;
3. Generality: The DQN algorithm does not rely on a specific network model, allowing it to be deployed across various heterogeneous network environments;
4. Offline learning: The DQN algorithm can be trained in a simulated environment before deployment. Once the strategy stabilizes, it can be implemented in real HWSNs application scenarios, thereby reducing the training time and cost in the actual network;
5. Action value estimation: The DQN algorithm optimizes Q-values to identify the optimal strategy, which is particularly suitable for dynamic decision-making in the context of information transmission energy efficiency optimization in HWSNs.

Despite its theoretical advantages, directly applying the DQN algorithm to optimize energy efficiency in information transmission within HWSNs still presents several challenges. These challenges include ensuring the robustness and generalization capabilities of the algorithm, accurately modeling the dynamic information transmission environment for DQN learning in HWSNs, and addressing the extensive data requirements and prolonged training times needed by the DNNs within the DQN algorithm to achieve a satisfactory performance. Furthermore, the DQN algorithm must be capable of making correct decisions in the rapidly changing information transmission scenarios typical of HWSNs. Therefore, this paper proposes adjustments and optimizations to the DQN algorithm to better accommodate these specific application scenarios.

This paper proposes a routing optimization protocol based on deep reinforcement learning, aiming to optimize data transmission decisions in HWSNs. The protocol comprehensively considers the combined effects of the distance between heterogeneous nodes, residual energy states, and the number of relays. Specifically, it first defines the coordinates and energy states of each node and cluster head as the current state. Then, utilizing the DQN algorithm, the protocol learns and selects the next-hop data transmission route based on a reward function until the data are successfully transmitted to the BS.

4. Improved DQN Algorithm

The DQN algorithm is one of the core algorithms in deep reinforcement learning, aiming to approximate the value function through representation learning. Unlike traditional Q-learning, the DQN leverages the powerful fitting and approximation capabilities of DNNs by taking the current state as input and outputting the corresponding Q-values for each possible action. Through training the DNN, the DQN algorithm gradually optimizes the Q-value predictions by minimizing the mean squared error between the predicted Q-values and the target Q-values. The core idea of the DQN is to enhance learning stability through experience replay and target networks. The experience replay mechanism stores the transitions encountered by the Agent while exploring the environment in a buffer, and during training, small batches of experience samples are randomly drawn for learning. This approach effectively reduces the correlation between samples, preventing drastic fluctuations in network parameter updates.

Therefore, deep reinforcement learning can be applied to solve decision-making problems with complex state and action spaces. In this paper, we design a scenario in HWSNs where super nodes with a high initial energy and computational capabilities act as Agents. These Agents learn optimal action strategies by interacting with the environment to maximize cumulative rewards. In practical applications, the DQN algorithm utilizes the fitting and approximation capabilities of DNNs to find optimal strategies, thereby handling high-dimensional state spaces and complex action spaces while exhibiting strong robustness.

4.1. Reward Function of HDQN Algorithm

In the data transmission phase, the BS receives the coordinates of all cluster heads. Subsequently, a DRL method is employed to determine the optimal data transmission route. In this scenario, the Agent aims to identify the best path by taking a series of actions in different states and receiving corresponding rewards to find the optimal data transmission route. The definitions of state s_t, action a_t, and reward r_t at the current time t are as follows:

1. State s_t: The state s_t includes the coordinates and remaining energy of heterogeneous ordinary nodes and cluster heads, the number of relay nodes involved in transmitting data from the cluster heads to the BS, and the header of the data packet.
2. Action a_t: The action a_t corresponds to the selection of the next-hop route.
3. Reward r_t: The reward function is defined as follows:

$$R_N = \omega_1 D(n) + \omega_2 E(n) - \omega_3 C(n) \tag{18}$$

where $D(n)$ represents the distance to the next-hop relay node. The goal is to select the node that is closest to both the current node and the BS as the next-hop node. In this paper, the proposed combination reward for distance and step size is given by

$$D(n) = \mu d_n + (1 - \mu) r_d \tag{19}$$

where d_n represents the distance between the source node M_{so} and the destination node M_{de} and $r_d = -\frac{d_{M_{so}B}}{d_{M_{de}B}}$ is the distance reward. $d_{M_{so}B} = \sqrt{{x_{M_{so}B}}^2 + {y_{M_{so}B}}^2}$ is the distance from source node M_{so} to BS. $d_{M_{de}B} = \sqrt{{x_{M_{de}B}}^2 + {y_{M_{de}B}}^2}$ is the distance from the destination node M_{de} to BS. μ and $1 - \mu$ represent the weights of d_n and r_n, respectively.

$E(n)$ reflects the remaining energy of the next-hop relay node, and nodes with higher energy are more likely to be selected. Super nodes typically have a stronger remaining energy level compared to other nodes, so the probability of being selected as relay nodes is also higher.

$C(n)$ represents the number of relay nodes in the data transmission route. The higher the value of $C(n)$, the lower the probability of choosing this transmission line. This parameter is designed to ensure that fewer relay nodes are involved in alternative data transmission lines, while also balancing the participation of other heterogeneous nodes in the transmission process, minimizing the overall energy consumption of HWSNs. Assuming N_{CH} is the number of all cluster heads, the number of relay node participants in a candidate data transmission line can be represented as $\sum_{i=1}^{N_{CH}} CH\_part_i$. ω_1, ω_2, and ω_3 are the weight ratios of $D(n)$, $E(n)$, and $C(n)$, respectively, with the aim of achieving energy balance and extending the network lifetime. In this paper, $\omega_1 = \omega_2 = \omega_3$ and $\omega_1 + \omega_2 + \omega_3 = 1$.

4.2. Design of HDQN Algorithm

To enhance the energy efficiency of HWSNs, this paper proposes the HDQN optimization algorithm for information transmission efficiency in such networks. The algorithm utilizes a short-distance, multi-hop information transmission path selection method based on the free-space energy model as much as possible. Assuming that in the HWSNs application scenario, the ordinary node $node_{or\_a}$ forwards the sensed data and packet header information to the cluster head $node_{CH-A}$. Then, it selects a super node as the Agent to exchange them with the environment. In addition, the cluster head $node_{CH-A}$ broadcasts the packet header information redefined by the HDQN algorithm (including reward value, remaining energy, and distance to neighboring nodes) to the neighboring cluster head $node_{CH-B}$ and the cluster head $node_{CH-C}$. If cluster head $node_{CH-C}$ does not receive the sensed data from the cluster head $node_{CH-A}$ next, it will discard the header information content. Once the corresponding cluster head $node_{CH-B}$ receives complete sensed data, it

continues to transmit the sensed data and packet header information to the next relay node. The flowchart of the HDQN algorithm is shown in Figure 5.

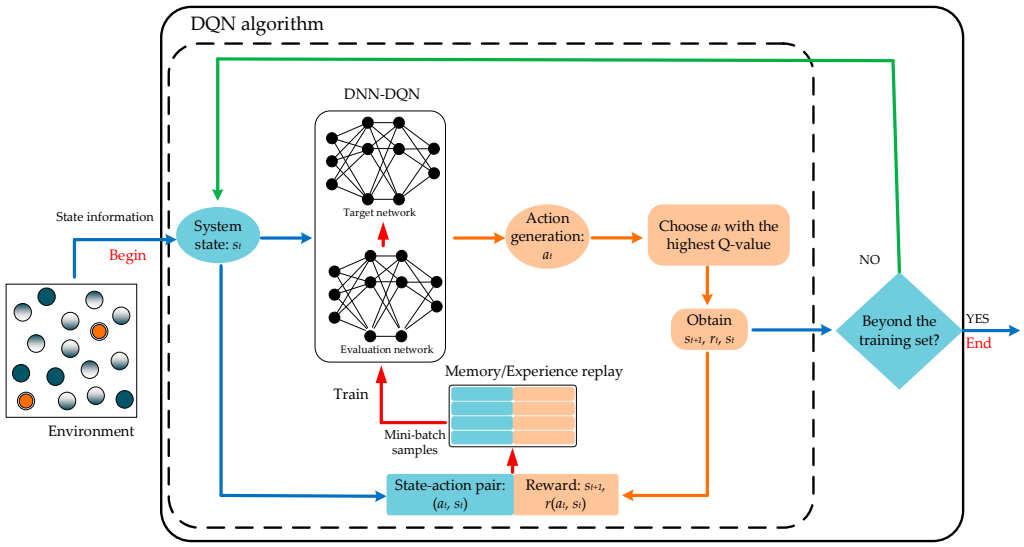

Figure 5. Flowchart of HDQN algorithm.

Figure 5 describes the flowchart of the HDQN algorithm, which addresses the information transmission path selection problem in HWSNs. The design of this algorithm has two main objectives. First, it aims to balance energy consumption across the entire HWSNs by assigning high-energy super nodes to act as Agents, cluster heads, and relay nodes, thereby handling more energy-intensive tasks. Second, it seeks to optimize information transmission by utilizing a short-distance, multi-hop method as much as possible.

The first step of the HDQN algorithm involves the super node acting as an Agent to perceive the state s_t of the HWSNs environment. This state information is then fed into the DQN evaluation network. The evaluation network generates corresponding reward $Q(s_t, a_t)$ values for all possible information transmission actions based on the current state s_t and selects the action a_t with the maximum Q-value for execution. Subsequently, the process of state, action, reward, and transition to the next state is stored as experience $\{s_t, a_t, r_t, s_{t+1}\}$ in the experience replay pool. To continuously train and improve the evaluation network, the HDQN algorithm randomly samples M mini-batches from the experience replay pool for training the evaluation network. Specifically, the loss function is computed according to Equation (20) to update the parameters of the evaluation network.

$$L_i(\delta_i) = \mathbb{E}_{s_t, a_t}\left[\left(r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}|\delta_{i-1}) - Q(s_t, a_t|\delta_i)\right)^2\right] \tag{20}$$

where δ represents the parameters of the DNN and i denotes the iteration index. Through this method, the HDQN algorithm can determine the optimal path for information transmission in HWSNs.

As shown in Figure 6, the HDQN algorithm improves the packet header structure by incorporating information relevant to DQN learning.

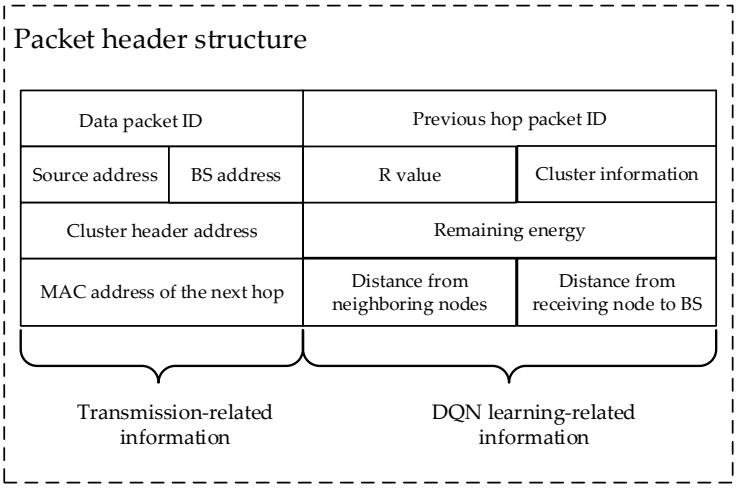

Figure 6. Improved HDQN algorithm packet header.

As illustrated in Figure 6, the original packet header structure contains the following key pieces of information relevant to data transmission:

Data packet ID: This field serves as a unique identifier for the packet, ensuring that each packet can be uniquely recognized and tracked within the network;

Source address: This indicates the network location of the sender node, specifying where the data packet originates;

Cluster header address: This represents the address of the cluster head node to which the source node belongs. The cluster head acts as a local controller or aggregator for data coming from nodes within its cluster;

BS address: This refers to the address of the BS where the packet is ultimately intended to be delivered. The BS serves as the final destination for the data in most cases, especially in hierarchical network structures.

This packet header structure is critical for ensuring that data packets are correctly routed through the network to their intended destinations, with clear identification of the source, the relevant cluster head, and the BS.

Additionally, to enhance the learning efficiency of the HDQN algorithm, the following DQN learning-related information has been innovatively included:

1. R value: This represents the cumulative expected reward value for performing a specific action in a given state, helping to evaluate the long-term benefits of actions taken by the Agent;
2. Remaining energy of current heterogeneous node: This indicates the remaining energy level of the current heterogeneous node, which is crucial for assessing its capability to perform tasks and transmit data;
3. Previous hop packet ID: This field denotes the packet ID and address of the previous hop's source node, aiding in tracking the packet's origin and its route through the network;
4. Distance to neighboring nodes: This represents the physical distance between the current node and its neighboring nodes, which is important for determining optimal routing paths and connectivity;
5. Distance from receiving node to BS: This indicates the distance from the current node to the BS, providing essential information for optimizing the transmission path and efficiently managing energy resources.

The improved packet header structure facilitates the efficient and real-time learning of the HDQN algorithm, thereby enhancing the overall energy efficiency of the HWSNs.

Algorithm 1 describes the pseudo-code for solving the information transmission path selection problem in HWSNs using the HDQN algorithm. The input to the algorithm is

the current state information of the HWSNs, and the output is the optimized information transmission path selection policy.

The process begins by initializing the parameters for the HDQN algorithm, the evaluation network, the target network, and the experience replay pool. Typically, the Agent selects the action with the highest Q-value. However, to ensure exploration of the environment in the current HWSNs, there is also a certain probability of randomly selecting an action. During execution, ordinary nodes directly transmit sensed information to the cluster head node. Subsequently, the Agent calculates the corresponding reward based on the reward function and updates it to the next state.

After the state update, the learning process begins. To obtain new network inputs and improve its learning efficiency, the HDQN algorithm transfers previously accumulated stored state from experience replay memory to the experience replay pool. It then randomly samples mini-batches from this pool and uses them to update the evaluation network, calculating the target Q-value for the current state. This entire process is iteratively performed across multiple training and learning cycles, ultimately yielding the optimized information transmission path policy for the HWSNs.

Algorithm 1. HDQN Protocol Pseudo-Code

Input: Node state information
Output: Routing policy
Initialize the evaluation network and target network parameters using parameter δ.
Initialize the experience replay memory D.
Set the super node as the Agent to interact with the environment.
Iterate over episodes $\{1, 2, \cdots, N^{\text{eps}}\}$:
Initialize state information s_t.
Iterate over time steps $t = \{1, 2, \cdots, T\}$.
Obtain the current state s_t.
Select action $a_{t1} = \text{argmax}Q(s_t, a_t)$ according to the ε-greedy strategy.
Randomly select action based on the $1 - \varepsilon$ strategy.
Ordinary nodes forward data packets to the higher-level node, and obtain the corresponding reward r_t and next state information s_{t+1} according to Equation (18).
Update the current state to the new state to acquire updated network input.
Store action $\{s_t, a_t, r_t, s_{t+1}\}$ in the experience replay memory.
Begin the learning process:
Randomly sample m mini-batch samples from the experience replay memory.
Update the evaluation network according to Equation (20).
Calculate the target Q-value $y_i = \begin{cases} r_j, \text{ If the data is successfully sent to BS} \\ r_j + \gamma \max_{a_{t+1}} \hat{Q}(\phi_{j+1}, a_{t+1}; \theta^-), \text{ other} \end{cases}$ for the current state.
Periodically update the target network.

4.3. Algorithm Performance Analysis

4.3.1. Time Complexity Analysis of the HDQN Algorithm

To discuss the time complexity of the DQN algorithm, it is necessary to first analyze the time complexity of its core DNN. This is because the DQN algorithm approximates the maximum cumulative reward through DNNs, so analyzing the time complexity of DNNs is crucial. The time complexity of DNNs is influenced by multiple key factors, including the number of layers (i.e., depth), the number of neurons contained in each layer, and the size of the convolution kernel. This paper assumes that the total number of layers in a DNN is $\mathbb{L}$, the edge length of the feature map output by the convolution kernel of the l-th layer is S_l, the size of the convolution kernel is M_l, the number of input channels is C_{l-1}, and the number of output channels is C_l. The time complexity of the DNN can be obtained as $O(\sum_{l \in L} S_l^2 \times M_l^2 \times C_{l-1} \times C_l)$.

As shown in the system model of Figure 2, there are N sensor nodes in the HWSNs. Assuming there are N_{CH} clusters and A_{total} actions in the DQN algorithm, the time complexity of the HDQN information transmission energy efficiency optimization algorithm is $O(\sum_{l \in L} (S_l^2 \times M_l^2 \times C_{l-1} \times C_l) \times A_{total} \times N \times N_{CH})$. It can be seen that the computational complexity of this algorithm has advantages compared to other exponentially increasing information transmission energy efficiency optimization algorithms.

4.3.2. Convergence Analysis of HDQN Algorithm

As shown in Equation (21), during the back-propagation training process of the DNN in the HDQN algorithm, the loss function L is calculated based on the immediate reward value r_{t+1}. Then the network parameters are updated through the iterative optimization of the objective function until convergence:

$$\theta_t \xrightarrow{\min\limits_{\theta} E_1(\theta_{t+1})} \theta_{t+1} \xrightarrow{\min\limits_{\theta} E_2(\theta_{t+2})} \cdots \xrightarrow{\min\limits_{\theta} E_k(\theta_{t+k})} \theta_{t+k} \tag{21}$$

5. Algorithm Simulation and Result Analysis

To more accurately evaluate the strengths and weaknesses of the HDQN algorithm proposed in this paper, Matlab R2022b was used as the simulation tool. The simulation experiment platform was configured with an Intel Core i7 3.6GHz processor and 16 GB of RAM. Assume sensor nodes are randomly distributed within a square area measuring 200 m by 200 m (denoted as: 200 m × 200 m), with the BS initially positioned at coordinates (100, 100). The HDQN algorithm is compared with five other algorithms—Energy-Efficient Hierarchical Cluster Routing (EEHCHR) [9], Two-Level Heterogeneous Gateway-Based Energy-Aware Multi-hop Routing (2L-HMGEAR) [10], a Network-Configurable Optimized Genetic Algorithm (NCOGA) [11], Distributed Energy-Efficient Clustering (DEEC) [12], and Stable Election Protocol (SEP) [13]—based on metrics such as average remaining energy, node survival, and data transmission volume. The comparative algorithms are as follows:

1. EEHCHR: A hierarchical cluster routing algorithm that focuses on energy efficiency and aims to reduce energy consumption in the network.
2. 2L-HMGEAR: Suitable for dual-layer HWSNs, optimizing multi-hop routing through energy-aware methods.
3. NCOGA: A network configuration optimization method based on genetic algorithms, designed to improve the performance of wireless sensor networks.
4. DEEC: This algorithm optimizes the LEACH algorithm by considering the remaining energy of nodes during cluster head election, thereby extending the network lifecycle.
5. SEP: This protocol sets different thresholds for HWSNs, enabling advanced nodes to take on more cluster head roles and thus improving the network stability.

The simulation experiment parameters for HWSNs are shown in Table 1.

Table 1. Table of parameters of HWSNs simulation experiment.

Simulation Parameter	Values
Sensor node monitoring area (m^2)	200 m × 200 m
BS locations (m)	(100, 100), (150, 150), (200, 200)
Number of wireless sensor nodes	100, 200, 300
Maximum round	3000
Energy-heterogeneous node type	super nodes, advanced nodes, ordinary nodes
Proportion of energy-heterogeneous nodes (super nodes/advanced nodes/ordinary nodes)	1:2:7
Normal energy-heterogeneous node initial energy (mJ)	300
Advanced energy heterogeneous node initial energy (mJ)	600

Table 1. *Cont.*

Simulation Parameter	Values
Super energy-heterogeneous node initial energy (mJ)	900
Unit transmission energy consumption E_{elec} (mJ/bit)	5×10^{-5}
Multipath fading loss coefficient E_{amp} (mJ/bit)	1.3×10^{-12}
Free-space loss coefficient E_{fs} (mJ/bit)	1×10^{-10}
Distance threshold d_{th} (m)	87.7
Energy per bit for information fusion E_{EDA} (mJ/bit)	5×10^{-6}

5.1. Simulation Experiment on the Node Survival Rate in HWSNs

After setting up the clustering, six algorithms are run with the same initial parameters to compare the node survival rate in HWSNs. As shown in Figure 7, the HDQN algorithm is compared with EEHCHR, 2L-HMGEAR, NCOGA, DEEC, and SEP. The aim is to observe how the node survival rate varies with the number of operational rounds, with the maximum number of rounds set to 1800 and a total of 100 nodes in the network. The NCOGA algorithm is the first to deplete the energy of an initial heterogeneous node at round 111. In contrast, the HDQN algorithm depletes its first heterogeneous node's energy at round 179, and the SEP algorithm experiences the depletion of its first node's energy at round 198. This indicates that SEP performs the best in terms of node survival rate during the early stages of operation, due to its specialized optimization design that accounts for the differences in initial energy levels among ordinary, advanced, and super nodes in HWSNs. This design effectively prolongs node survival, especially in the initial phases.

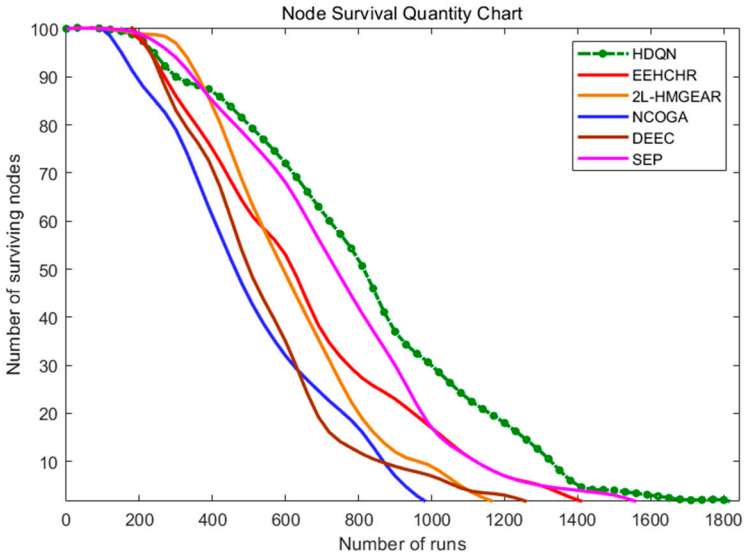

Figure 7. Comparison of simulation experiments on the number of surviving nodes in HWSNs.

However, as the number of operational rounds increases, from the depletion of the 25th node to the depletion of all nodes, the HDQN algorithm consistently ranks first among all six algorithms. This superior performance is attributed to the HDQN's ability to allocate high-energy roles, such as Agents, cluster heads, and communication relays, to nodes with more initial energy. Additionally, the HDQN uses DNNs to intelligently manage and optimize information transmission paths. Thus, the HDQN algorithm demonstrates a significant advantage in terms of node survival rate in heterogeneous networks, as the number of surviving nodes gradually decreases over time.

Table 2 provides specific data from the simulation experiment comparing the node survival rates in HWSNs. As the number of operational rounds increases, a higher number of surviving heterogeneous nodes indicates the better optimization of information transmission efficiency in the HWSNs. A sufficient number of surviving heterogeneous nodes ensures the coverage of the monitoring area and the completion of their assigned tasks. According to Table 2, the NCOGA algorithm was the first to deplete the energy of a node, while the 2L-HMGEAR and SEP algorithms saw the depletion of the first heterogeneous node's energy in the 195th and 198th rounds, respectively. As the number of rounds continues to increase, the HDQN algorithm leads in the rounds at which the energy of the 25th, 50th, 75th, and 100th nodes is depleted compared to the second-place SEP algorithm. Specifically, these rounds were 44, 79, 128, and 131, with improvements in node survival time proportions of 8.58%, 10.81%, 13.78%, and 7.65%, respectively. This indicates that the HDQN algorithm consistently extends the survival time of heterogeneous sensor nodes, thereby enhancing the efficiency and reliability of HWSNs.

Table 2. Simulation experiment data on number of surviving nodes in HWSNs.

Algorithm	The 1st Node's Energy Consumption Rounds	The 25th Node's Energy Consumption Rounds	The 50th Node's Energy Consumption Rounds	The 75th Node's Energy Consumption Rounds	The 100th Node's Energy Consumption Rounds
EEHCHR	183	398	644	823	1458
2L-HMGEAR	195	437	601	754	1305
NCOGA	111	319	455	696	1006
DEEC	187	357	506	653	1431
SEP	198	513	731	929	1713
HDQN	179	557	810	1057	1844

5.2. Simulation Experiment of Average Remaining Energy of Nodes in HWSNs

As presented in Figure 8, after running for the same number of rounds, a comparison of the average remaining energy of the nodes among the six algorithms, including HDQN and five other algorithms, is conducted. This comparison focuses on how the average remaining energy of the heterogeneous nodes changes with the number of rounds. The maximum number of rounds is set at 1600. At round 1006, the average remaining energy of nodes running the NCOGA algorithm drops to 0 mJ. For the algorithms 2L-HMGEAR, DEEC, EEHCHR, SEP, and HDQN, the rounds at which the average remaining energy of nodes first drops to 0 mJ are 1305, 1431, 1458, 1713, and 1844, respectively. The experimental results indicate that the HDQN algorithm, by employing deep reinforcement learning to select information transmission paths, significantly enhances the efficiency of energy utilization in heterogeneous nodes. This leads to a more balanced energy consumption across nodes. Consequently, the HDQN algorithm maintains a higher average remaining energy of nodes over the same number of rounds compared to the other algorithms.

The simulation experiment data presented in Table 3 compare the average remaining energy of the nodes in HWSNs using different algorithms. As the number of operational rounds increases, a higher average remaining energy of heterogeneous nodes signifies the better optimization of energy efficiency, allowing the network to perform tasks such as information collection, transmission, and processing more efficiently over an extended period without interruptions due to energy depletion. From Table 3, it is observed that after 500 rounds, the SEP algorithm demonstrates the best performance in terms of the average remaining energy of its heterogeneous nodes, with the proposed HDQN algorithm coming in second. However, as the number of rounds progresses, the advantages of the HDQN algorithm become increasingly apparent. After 600 rounds, the HDQN algorithm surpasses the other five algorithms, including SEP, in terms of the average remaining energy of its nodes.

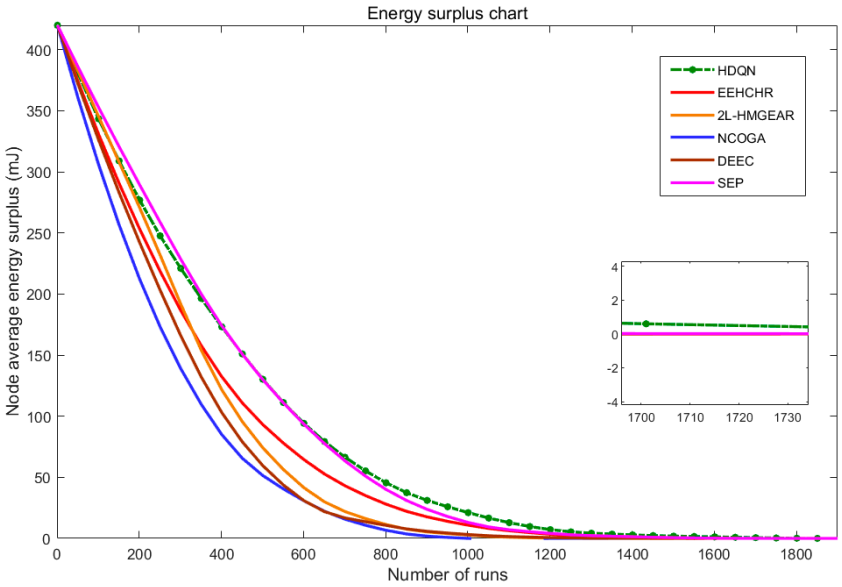

Figure 8. Comparison of average remaining energy of nodes in HWSNs.

Table 3. Simulation data of average remaining energy of nodes in HWSNs (unit: mJ).

Algorithm	Average Remaining Energy at 500 Rounds	Average Remaining Energy at 600 Rounds	Average Remaining Energy at 700 Rounds	Average Remaining Energy at 800 Rounds	Average Remaining Energy at 900 Rounds
EEHCHR	94.47	64.754	41.895	28.111	18.696
2L-HMGEAR	74.1528	41.648	21.203	11.391	5.909
NCOGA	52.647	30.94	16.57	6.72	1.631
DEEC	58.44	31.002	16.912	10.675	6.478
SEP	131.190	93.780	63.691	40.212	24.038
HDQN	128.635	94.514	66.263	45.641	31.338

This improvement can be attributed to the HDQN algorithm's unique approach of utilizing the real-time network status sensed by the Agent and employing decision-making for information transmission paths using an improved DQN algorithm. This mechanism effectively reduces energy consumption in heterogeneous nodes and enhances the network's sustained operation capability. Although the HDQN algorithm's average remaining energy was slightly behind the SEP algorithm by 1.98% in the early stages (before 600 rounds), it began to lead the other five algorithms from the 600th round onwards. By this point, the HDQN algorithm showed improvements of 0.78%, 4.04%, 13.50%, and 15.37% over the 2L-HMGEAR, DEEC, EEHCHR, and SEP algorithms, respectively. This indicates that the HDQN algorithm continuously enhances the energy efficiency of HWSNs, with its optimization effects becoming more pronounced in the later stages of network operation. In summary, the HDQN algorithm's superior performance, especially during the later stages of the simulation, highlights its effectiveness in extending the operational lifetime and reliability of HWSNs through improved energy management and utilization.

5.3. Simulation Experiment of Total Energy Consumption in HWSNs

As shown in Figure 9, six algorithms were executed with the same initial parameters for a maximum of 1800 rounds to observe how the total energy consumption of heterogeneous wireless sensors changes with the number of rounds. By the time 900 rounds were completed, the energy consumption of the DEEC, NCOGA, EEHCHR, and 2L-HMGEAR algorithms had all exceeded 40,000 mJ. The second-best performing SEP algorithm also reached an energy consumption of 40,000 mJ after approximately 1000 rounds. It can be observed that as the number of operational rounds increases, the total energy consumption

of the networks using the DEEC, NCOGA, 2L-HMGEAR, EEHCHR, and SEP algorithms rises sharply. This can be attributed to the fact that the proposed HDQN algorithm, designed to optimize energy efficiency in HWSNs, comprehensively considers the energy variability among different heterogeneous nodes. Additionally, it utilizes improved deep reinforcement learning to optimize the selection of information transmission paths in real time.

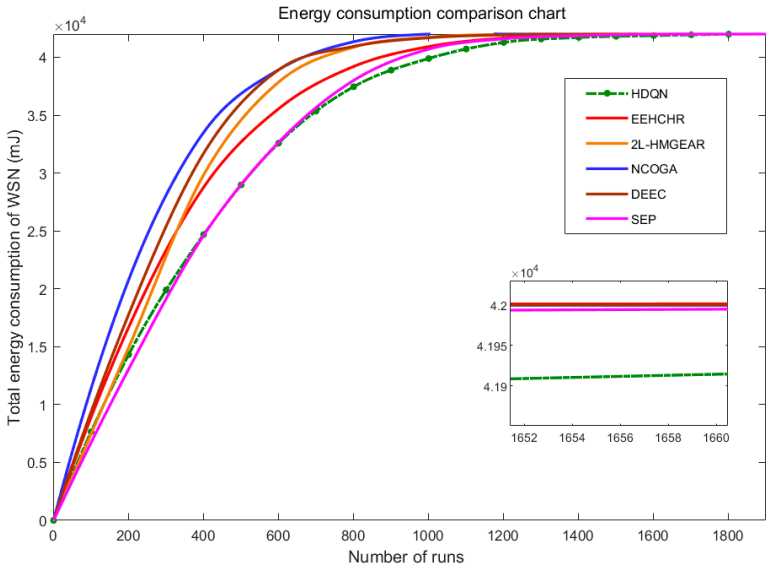

Figure 9. Comparison of total energy consumption in HWSNs.

The specific data from the simulation experiment comparing the total energy consumption of HWSNs are presented in Table 4. A lower total energy consumption indicates better energy efficiency optimization, which can enhance network reliability and reduce the costs associated with maintaining and replacing nodes. From Table 4, it can be observed that after 600 rounds, the total energy consumption of the network using the proposed HDQN algorithm is 32,548.6 mJ, which is the lowest among the six compared algorithms. Specifically, the total energy consumption of nodes at rounds 600, 700, 800, and 900 using the HDQN algorithm is lower than that of the second-best SEP algorithm by 73.4 mJ, 257.3 mJ, 543.6 mJ, and 729.4 mJ, respectively. The reduction in total energy consumption is 0.23%, 0.73%, 1.45%, and 1.88% correspondingly. This demonstrates the excellent performance of the HDQN algorithm in terms of total energy consumption in HWSNs.

Table 4. Simulation data of total energy consumption in HWSNs (unit: mJ).

Algorithm	Total Energy Consumption at 500 Rounds	Total Energy Consumption at 600 Rounds	Total Energy Consumption at 700 Rounds	Total Energy Consumption at 800 Rounds	Total Energy Consumption at 900 Rounds
EEHCHR	32,552.4	35,524.5	37,810.5	39,188.8	40,130.4
2L-HMGEAR	34,584.7	37,835.1	39,879.7	40,860.8	41,409.1
NCOGA	36,735.3	38,905.4	40,342.9	41,327.8	41,836.9
DEEC	36,155.6	38,899.8	40,308.8	40,932.4	41,352.2
SEP	28,880.9	32,622.0	35,630.9	37,978.7	39,596.1
HDQN	29,136.5	32,548.6	35,373.6	37,435.1	38,866.7

5.4. Simulation Experiment of Energy Consumption with Varying BS Positions in HWSNs

The comparative experiment of BS position changes in HWSNs is designed to verify the robustness of the energy efficiency optimization algorithms. In HWSNs, the positions of BSs may vary depending on different tasks. Under the same number of iterations, if the BS position changes, the proposed algorithm should still be able to operate efficiently and

consume less energy, demonstrating its robustness in energy efficiency optimization across various application scenarios of HWSNs. As shown in Figure 10, the HDQN algorithm is compared with five other algorithms to analyze the impact of BS position changes on energy efficiency optimization algorithms in HWSNs. The BS positions are set to (100, 100), (150, 150), and (200, 200). In these three BS positions, all six algorithms run for 700 rounds. The DEEC algorithm exhibits the highest total energy consumption, while the total energy consumption of the 2L-HMGEAR and NCOGA algorithms are similar, increasing as the BS position changes from (150, 150) to (200, 200). The SEP algorithm and the proposed HDQN algorithm show the best performance in terms of total network energy consumption, with the HDQN algorithm maintaining the lowest energy consumption across all three BS positions. The simulation results indicate that the HDQN algorithm demonstrates good robustness in energy efficiency optimization in various application scenarios with different BS positions.

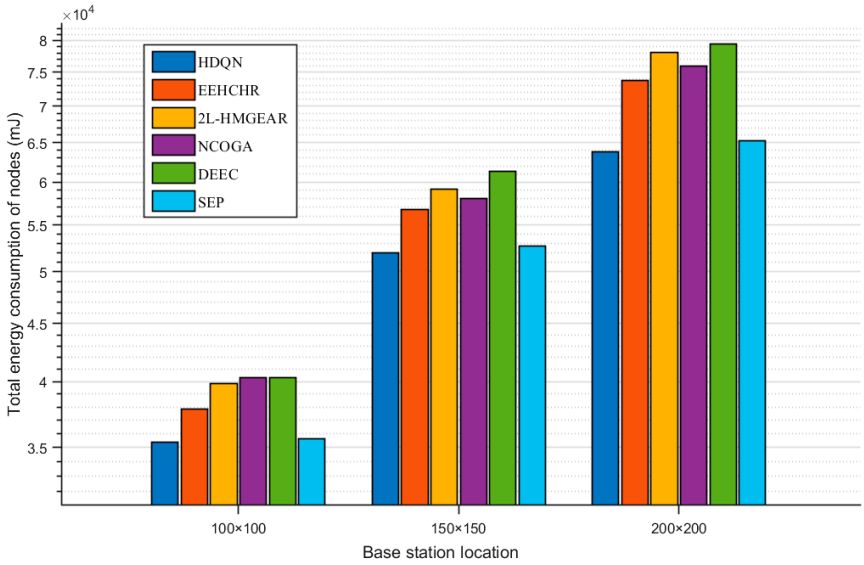

Figure 10. Comparative chart of total energy consumption in HWSNs with varying BS positions.

Table 5 presents the network energy consumption of various energy efficiency optimization algorithms after 700 rounds of operation in HWSNs with different BS positions. The SEP algorithm demonstrates relatively low energy consumption at all BS positions, with a particularly noteworthy performance at the BS position of (100, 100), where it consumes only 35,630.9 mJ, ranking second only to the proposed HDQN algorithm. The HDQN algorithm consistently exhibits the lowest energy consumption across the three BS positions. Specifically, at BS positions of (100, 100) and (150, 150), the HDQN algorithm reduces energy consumption by 257.3 mJ and 709.2 mJ, respectively, compared to the second-ranked SEP algorithm, representing reductions of 0.72% and 1.35%. At the BS position of (200, 200), the performance difference is most pronounced, with the HDQN algorithm achieving a 1392.6 mJ reduction in energy consumption compared to the SEP algorithm, representing a 2.13% reduction. This indicates that the HDQN algorithm employs a specialized energy efficiency optimization strategy for HWSNs and demonstrates good robustness across various application scenarios.

Table 5. Simulation data of energy consumption with varying BS positions in HWSNs (unit: mJ).

Algorithm	Energy Consumption After 700 Rounds (mJ)		
	BS Location (100, 100)	BS Location (150, 150)	BS Location (200, 200)
EEHCHR	37,810.5	56,742.1	73,782.6
2L-HMGEAR	39,879.7	59,101.76	78,013.3
NCOGA	40,342.9	57,986	75,961.7
DEEC	40,308.8	61,311.8	79,485.2
SEP	35,630.9	52,684.2	65,241.9
HDQN	35,373.6	51,975	63,849.3

5.5. Simulation Experiment of Energy Consumption with Varying Node Numbers in HWSNs

In various scenarios of HWSNs operations, the number of nodes typically changes according to different application requirements. If the proposed algorithm exhibits good energy efficiency optimization performance across different node counts in HWSNs, it indicates the strong adaptability of the algorithm. As illustrated in Figure 11, to validate the performance of the HDQN algorithm in energy efficiency optimization with varying node numbers, it is compared with five other algorithms. The total network energy consumption of various algorithms is analyzed as the number of nodes changes. The number of nodes in the HWSNs is set to 100, 200, and 300, with all six algorithms running for 700 rounds under identical initial parameters. The proposed HDQN algorithm consistently ranks among the top in terms of total remaining energy across all three node counts. This observation indicates that the HDQN algorithm effectively reduces network energy consumption in scenarios with varying numbers of HWSNs nodes, demonstrating its robust adaptability and energy efficiency optimization capabilities.

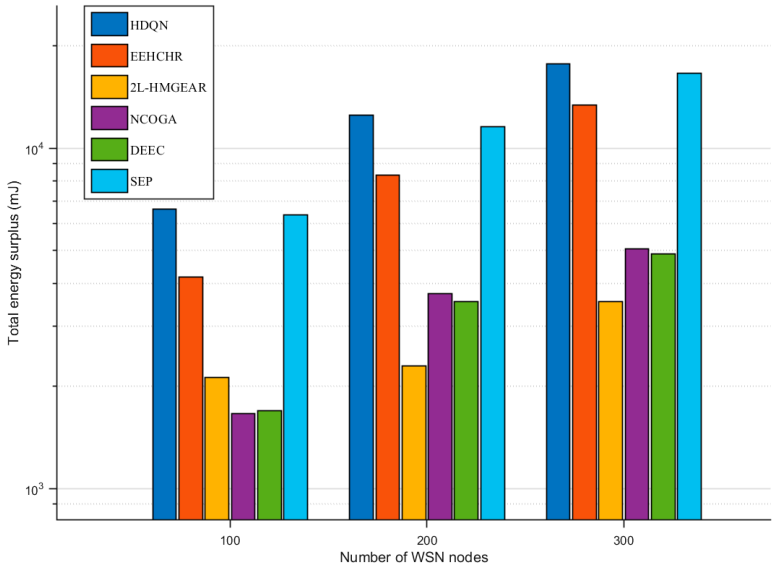

Figure 11. Simulation experiment of remaining energy with varying node numbers in HWSNs.

Table 6 presents the total remaining energy of various energy efficiency optimization algorithms after 700 rounds of operation in scenarios with different numbers of heterogeneous sensor nodes. As depicted in Table 6, the HDQN algorithm exhibits superior performance compared to the second-ranked algorithm across node counts of 100, 200, and 300, achieving energy consumption reductions of 257.2 mJ, 944.7 mJ, and 1071.3 mJ, respectively. The corresponding percentage improvements stand at 4.04%, 8.19%, and 6.46%. This suggests that the HDQN algorithm effectively optimizes energy efficiency in

HWSNs across varying numbers of sensor nodes, demonstrating its robust performance and adaptability to different deployment scenarios.

Table 6. Simulation data of energy consumption with varying node numbers in HWSNs (unit: mJ).

Algorithm	Energy Consumption After 700 Rounds (mJ)		
	For 100 Nodes	For 200 Nodes	For 300 Nodes
EEHCHR	4189.5	8314.2	13,354.8
2L-HMGEAR	2120.3	2296.4	3534.4
NCOGA	1657	3742.1	5056
DEEC	1691.2	3539.5	4897.7
SEP	6369.1	11,541.8	16,596
HDQN	6626.3	12,486.5	17,667.3

5.6. Simulation Experiment of Transmission Counts with Free-Space Energy Model in HWSNs

According to the classic energy equation for wireless sensor networks, under the same node and operation round conditions, a higher number of transmissions using the free-space energy model indicates a more effective energy efficiency optimization strategy for HWSNs. As depicted in Figure 12, the HDQN algorithm is compared with five other algorithms (DEEC, 2L-HMGEAR, NCOGA, SEP, and EEHCHR) in terms of the number of transmissions using the free-space energy model across heterogeneous node counts of 100, 200, and 300. This comparison examines how the number of transmissions using the free-space energy model varies with an increasing number of nodes for these six algorithms. In the scenario with 300 nodes and identical initial conditions, the DEEC algorithm exhibits the lowest number of transmissions using the free-space energy model. The 2L-HMGEAR, NCOGA, and SEP algorithms show similar transmission counts, which are slightly lower than those of the EEHCHR and HDQN algorithms. Notably, the number of transmissions using the free-space energy model for all six algorithms increases steadily with the number of nodes, which is consistent with the expectation of increased communication activity as the network size grows. Importantly, the proposed HDQN algorithm consistently ranks first in the number of transmissions using the free-space energy model across different node counts (100, 200, and 300 nodes). This indicates that the HDQN facilitates more transmissions under the same energy conditions, thereby demonstrating superior energy efficiency optimization. This performance underscores that the HDQN algorithm enhances energy efficiency optimization in various application scenarios of HWSNs and exhibits robust performance across different network sizes.

Table 7 illustrates variations in the number of transmissions using the free-space energy model across different scenarios, each involving varying numbers of heterogeneous nodes. Notably, the EEHCHR algorithm exhibits a relatively higher number of transmissions under the free-space energy model in all three scenarios, reaching 257,446 transmissions when the scenario includes 300 heterogeneous nodes. This observation suggests that the EEHCHR algorithm prioritizes direct information transmission between nodes that are spatially proximate to each other, thereby minimizing energy expenditure and enhancing data transmission efficiency. In contrast, the proposed HDQN algorithm achieves the highest number of transmissions using the free-space energy model across different node counts, with a remarkable 260,918 transmissions when 300 nodes are present. This represents a substantial 3472 transmissions more than the second-best EEHCHR algorithm, corresponding to a 1.35% improvement. These findings indicate that the HDQN algorithm not only effectively enhances information transmission efficiency in HWSNs but also demonstrates a robust adaptability across diverse scenarios characterized by varying numbers of heterogeneous nodes.

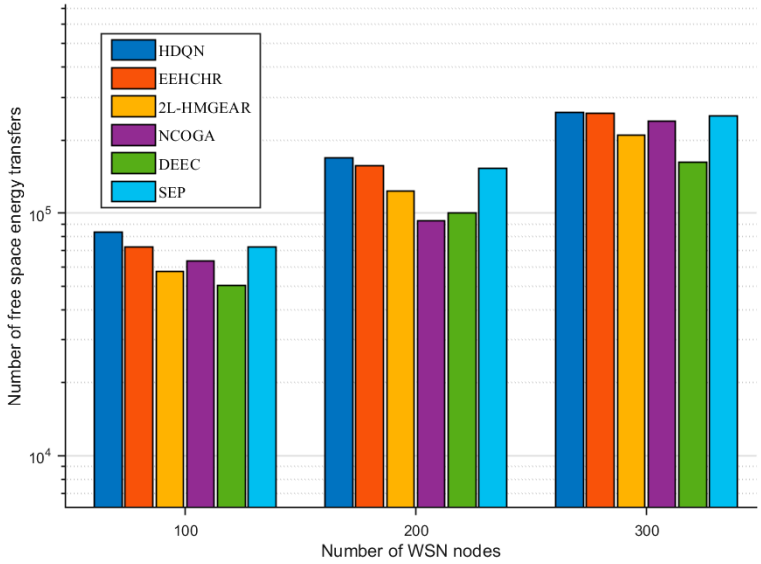

Figure 12. Simulation of transmission counts with free-space energy model in HWSNs.

Table 7. HWSNs free-space energy transmission frequency simulation experiment data.

Algorithm	The Number of Times Energy is Transferred Using Free Space		
	For 100 Nodes	For 200 Nodes	For 300 Nodes
EEHCHR	72,360	156,848	257,446
2L-HMGEAR	57,564	123,420	210,091
NCOGA	63,475	93,090	238,809
DEEC	50,216	100,253	162,343
SEP	72,402	153,372	253,011
HDQN	83,503	169,091	260,918

6. Conclusions

This paper tackles the challenge of enhancing the overall energy efficiency in HWSNs through the exploitation of energy disparities among nodes. We introduce a DQN-based information transmission path selection algorithm, termed HDQN, which leverages the energy heterogeneity of the network and environmental characteristics to make real-time decisions on information transmission paths, thus improving the overall network efficiency. The HDQN algorithm employs the DQN framework to manage the intricate action space typical of network environments, focusing on optimizing energy efficiency. Our design of the reward function incorporates elements such as node distance, remaining node energy, and the number of relay nodes to balance the involvement of various heterogeneous nodes in the information transmission process and minimize overall energy consumption in the WSNs. The loss function within the DQN network utilizes mean squared error to compare the Q-value of the action taken in the current state with the anticipated cumulative reward, enabling a more precise estimation of cumulative rewards and enhancing decision-making strategies. Furthermore, we implement an improved packet header structure as input for network state information, facilitating the selection of appropriate data transmission optimization routing strategies. In particular, this packet header information is broadcasted to neighboring cluster head nodes, which subsequently propagate it to designated relay nodes until reaching the BS. This approach substantially boosts the overall energy efficiency of data transmission in HWSNs. Through simulation experiments comparing the average remaining node energy, number of surviving nodes, amount of information transmitted, total network remaining energy, and network robustness, our HDQN algorithm is evaluated against five other algorithms—EEHCHR, 2L-HMGEAR, NCOGA, DEEC, and SEP. The

results substantiate that the HDQN algorithm markedly advances energy efficiency and prolongs the network's lifetime in complex HWSNs environments.

Author Contributions: Methodology, Y.S. and X.H.; investigation, W.Z.; writing—original draft preparation, Y.S.; writing—review and editing, Z.L.; supervision, Z.L., Y.S., Z.L., X.H. and W.Z.; conceptualization, K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Key Science and Technology Plan Project of Zigong City (Zigong Medical Big Data and Artificial Intelligence Research Institute) (2022ZCYGY16), by the Teaching Reform Research Project of the Sichuan University of Science and Engineering (JG-24025), by the Sichuan University of Science and Engineering Graduate Education and Teaching Reform Research Project (JG202403), by the Third Phase of the Ministry of Education's Supply and Demand Docking Employment Education Project (2023122021680) in 2024, by the Second Batch of the Ministry of Education's Industry University Cooperation Collaborative Education Project (202102123021), by the Opening Project of the Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationization and the Internet of Things (2022WYY02), and by the Talent Introduction Project of the Sichuan University of Science and Engineering under Grant (2020RC22).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, M.; Zhang, S.; Cao, Y.; Xu, S. NMSFRA: Heterogeneous routing protocol for balanced energy consumption in mobile wireless sensor network. *Ad Hoc Netw.* **2023**, *145*, 103176. [CrossRef]
2. Kalra, H. The Heterogeneous wireless sensor networks: An energy-efficient cooperative routing scheme. In Proceedings of the 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Bangalore, India, 29–31 December 2023; pp. 1–7.
3. Karimi-Bidhendi, S.; Guo, J.; Jafarkhani, H. Energy-efficient deployment in static and mobile heterogeneous multi-hop wireless sensor networks. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 4973–4988. [CrossRef]
4. Jha, V.; Sharma, R. An energy efficient weighted clustering algorithm in heterogeneous wireless sensor networks. *J. Supercomput.* **2022**, *78*, 14266–14293. [CrossRef]
5. Rawat, P.; Rawat, G.S.; Rawat, H.; Chauhan, S. Energy-efficient cluster-based routing protocol for heterogeneous wireless sensor network. *Ann. Telecommun.* **2024**. [CrossRef]
6. Dutt, S.; Agrawal, S.; Vig, R. Cluster-head restricted energy efficient protocol (CREEP) for routing in heterogeneous wireless sensor networks. *Wirel. Pers. Commun.* **2018**, *100*, 1477–1497. [CrossRef]
7. Mateen, A.; Ahad, A.; Zia, S.; Shayea, I.; Ali, S. Energy-efficient routing to prevent void holes in heterogeneous 5G wireless sensor network using game theory. In Proceedings of the 2023 International Conference on Smart Computing and Application (ICSCA), Hail, Saudi Arabia, 5–6 February 2023; pp. 1–6.
8. Hung, L.-L.; Leu, F.-Y.; Tsai, K.-L.; Ko, C.-Y. Energy-efficient cooperative routing scheme for heterogeneous wireless sensor networks. *IEEE Access* **2020**, *8*, 56321–56332. [CrossRef]
9. Panchal, A.; Singh, R.K. EEHCHR: Energy efficient hybrid clustering and hierarchical routing for wireless sensor networks. *Ad Hoc Netw.* **2021**, *123*, 102692. [CrossRef]
10. Benelhouri, A.; Idrissi-Saba, H.; Antari, J. An improved gateway-based energy-aware multi-hop routing protocol for enhancing lifetime and throughput in heterogeneous WSNs. *Simul. Model. Pract. Theory Int. J. Fed. Eur. Simul. Soc.* **2022**, *116*, 102471. [CrossRef]
11. Sahoo, B.M.; Pandey, H.M.; Amgoth, T. A genetic algorithm inspired optimized cluster head selection method in wireless sensor networks. *Swarm Evol. Comput.* **2022**, *75*, 101151. [CrossRef]
12. Qing, L.; Zhu, Q.; Wang, M. Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. *Comput. Commun.* **2006**, *29*, 2230–2237. [CrossRef]
13. Smaragdakis, G.; Matta, I.; Bestavros, A. *SEP: A Stable Election Protocol for Clustered Heterogeneous Wireless Sensor Networks*; Boston University Computer Science Department: Boston, MA, USA, 2004.