Received 5 October 2023, accepted 14 October 2023, date of publication 18 October 2023, date of current version 25 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3325724

RESEARCH ARTICLE

An Efficient Method for Optimal Allocation of Resources in LPWAN Using Hybrid Coati-Energy Valley Optimization Algorithm Based on Reinforcement Learning

M. RAJESWARA RAO AND S. SUNDAR
School of Electronics Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India
Corresponding author: S. Sundar (sundar.s@vit.ac.in)

ABSTRACT In Wide Area Networks (WANs), resource allocation is essential since computer networks manage complex optimization difficulties. The goal of optimal resource allocation is to improve total computing productivity. Distributing numerous specialized data and communication technology services inside a WAN can be challenging due to a wide range of application demands. A system to provide enhanced data transfer rate with minimum power transmission in a Low Power Wide Area Network (LPWAN) with the utilization of reinforcement learning technique is developed in this work. Here, the reinforcement learning method is used to determine the parameters involved for minimizing the transmission power. The resources in the networks, like the channel, spreading factor, and transmission power in the LPWAN is optimized. These parameters optimization is aided with the help of the newly developed Hybrid Coati with Energy Valley Optimization Algorithm (HC-EVOA). A large number of reinforcement learning agents are generated that match the terminal hubs in the server of the LPWAN. Once the reinforcement agent is generated, then the optimized transmission parameter is given to these terminal hubs of the network. The optimization is carried out to enhance the throughput and reduce the energy consumption rate by the equipment in the network.Simulations are conducted on the developed model to prove the system's effectiveness. Based on the analysis, at transmission power of 8, the developed HC-EVOA-based optimal resource allocation system obtains energy consumption as 12.86% lower than GSO, 19.29% lower than ROA, 2.76% lower than COA, and 2.81% lower than EVO.

INDEX TERMS Optimal resource allocation, low-power wide area network, hybrid coati with energy valley optimization algorithm, reinforcement learning agent, spreading factor, transmission power, energy consumption rate.

I. INTRODUCTION

The fast growth of the Internet over the past decade has significantly provided extensive contributions to human society around the world [1]. One of the key problems in wireless networks are optimal for resource allocation, and hence numerous methods like standardization and game theory have historically been used to overcome this problem [2]. It is more important to note the struggle between the growing functional needs of network users and the inappropriate distribution of fundamental network resources in Internet infrastructures [3]. To deliver an enormous amount of network resources quickly to support their operation, WAN necessitates an impressive and strong basic network design [4]. In the standard WAN, there are multiple approaches, such as heuristic strategies, discontinuous optimization-assisted methods, and methods based on convex optimization are used for distributing resources [5]. Nevertheless, they must possess an in-depth

The associate editor coordinating the review of this manuscript and approving it for publication was Eyuphan Bulut.

understanding of Channel State Information (CSI), which may not be feasible in dynamic wireless environments because of the significant latency associated with feedback channels [6]. Multi-objective problems are not solved by conventional techniques. The tradeoff between spectral effectiveness, energy consumption, and equity emerged as one of the main difficulties in the area of resource allocation for future WANs.

There are a lot of linked devices present within the numerous IoT applications, but they produce only a tiny amount of data. With the help of this IoT function, enormous relationships and an LPWAN are made feasible [7]. LPWAN techniques are utilized to meet the needs of IoT applications, and they balance energy usage, protection, and data rates compared to classic wireless techniques [8]. LPWAN methods run at minimal data rates to offer long-distance transmission with less energy consumption makes them better suited for delay-tolerant applications with limited data requirements [9]. Long-range (LoRa) is one of the more exciting LPWAN technologies currently in use, and it receives a lot of attention from both the business and academic worlds [10]. Chirp spread spectrum technology employs a variety of Spreading Factors (SF) with little power consumption and is considered the foundation of LoRa's success [11]. The physical layer operation of LoRa devices has been suggested in the LoRaWAN technique to enhance the lifespan of batteries and network performance. Most modern LPWAN technologies, including LoRa, operate in the unlicensed spectrum that can be extensively accessible for Internet of Things applications. LoRaWAN is going to be interference-limited rather than noise-limited under hyper-dense deployment scenarios. Coexistence difficulties become more crucial for LPWAN. The server will adjust SF and rise or fall the transmit power according to the Signal-to-Noise Ratio (SNR) to improve transmission speeds, airtime, and energy consumption in the network. The transmit power is adjusted at each step of the centralized method, which results in relatively low resource efficiency, especially in urban-scale IoT applications that require a large number of connected devices. To prevent channel conflicts and boost the efficiency of resources, an intelligent resource allocation system is necessary [12].

Deep reinforcement learning and other reinforcement learning techniques have recently been employed for allocating resources in wireless networks [13]. A new machine learning algorithm called the Deep reinforcement learning algorithm combines the deep learning algorithm and the reinforcement learning algorithm [14]. The selection issue in space with high dimensions and space of states is primarily solved by the Deep reinforcement learning algorithm [15]. Artificially formed neural networks are mostly used in the deep learning method to handle high-dimensional issues in decision-making [16]. The key objective of reinforcement learning is to employ an agent to know about the decision-making protocol to communicate with its surroundings and receive the most beneficial reward [17]. By the interactions with a learning environment, such as the number of users, discoloration, and quality of service needs. Because of the above interaction, a reinforcement agent discovers the ideal policy for resource allocation in Long Range Wide Area Networks (LoRaWAN). An improved deep reinforcement learning approach is suggested to eliminate transmission power while keeping reliability, latency, and transmission rate restrictions. However, identifying of optimal policies in the surrounding era is the toughest job. Additionally, several theoretical principles must be quickly reinforced in the early stages. Hence, a novel optimal resource allocation system in LPWAN is developed in this work.

The important contributions of reinforcement learning-based resource allocation system are portrayed in underneath part.

- To build up the reinforcement learning-based optimal resource allocation process for allocating resources to the terminal nodes of the LPWAN. So it is easy to achieve a wide range of communication with low power and cost.
- To enrich the performance of the reinforcement learning-based optimal resource allocation process, HC-EVOA is developed, and it is utilized for optimally selecting the spreading factor, transmission power, and channel for conducting the optimal resource allocation process.
- To allocate the resources to the terminal nodes of the LPWAN, an HC-EVOA, is adopted to optimizing the attributes. It is used to lower the Signal to Interference plus Noise Ratio (SINR) during the resource allocation process by optimizing the variable like spreading factor, transmission power and channel.
- The outcome of the HC-EVOA-based resource allocation system is compared with the traditional resource allocation system to know the performance of the HC-EVOA-based resource allocation system in terms of various measures.

The invented HC-EVOA-based resource allocation system is categorized as follows. Section II details the traditional resource allocation system and its features and challenges. In section III, the network model of low-power wide area network, and its problem regarding performance and data rate is presented briefly. The hybrid coati with energy valley optimization algorithm for selecting the optimal reinforcement learning parameters for resource allocation in LPWAN is presented in section IV. In section V, efficient reinforcement learning-based optimal resource allocation in LPWAN using a hybrid optimization algorithm is given. The result and discussion of the HC-EVOA-based resource allocation system are depicted in section VI. The conclusion of the HC-EVOA-based resource allocation system is portrayed in section VII.

II. LITERATURE SURVEY

Many studies have explored diverse research approaches to enhance resource allocation in LoRaWAN networks, focusing on optimizing transmission power, energy efficiency, and network performance. For a more comprehensive understanding of these research endeavours, the following sections concisely describe each study discussed:

A. DISCUSSION ON RESOURCE ALLOCATION MODELS

In 2020, Park et al. [18] developed a resource allocation system for LoRaWAN using reinforcement learning techniques. Here, the transmission energy was lowered by considering the reinforcement learning system used to identify the attributes, so that the transmission power was greatly lowered. The network resources like spreading factor, channel, and transmission power were optimized with the help of deep reinforcement learning techniques to overcome the issues involved in the LoRaWAN. In WAN, the maximized transmission energy is given to all network terminal nodes using deep reinforcement learning techniques. The experimental output showed that the performance of the invented system was superior in terms of throughput.

In 2023, Gava et al. [19] recommended an effective resource distribution system for LoRaWAN using a minimum-cost spanning tree algorithm. In this work, the computational cost and the energy utilization during the resource allocation process were greatly lowered by the "minimum-cost spanning tree algorithm" and "Variable Neighborhood Search (VNS)". The VNS was also used to identify the repeaters' location in the LoRaWAN. Modifying attributes like transmission power, SF, and bandwidth lowered the total energy utilization and time required to gather the data. The link in the middle of time on air and the Ebit was determined by the SF attribute. The SF attribute determined the link in the middle of time on air and the Ebit. Fixing a small SF value for the network with the minimum number of devices was recommended; therefore, the Ebit value was greatly lowered. Furthermore, the value of SF was taken as large for the extension process. Stimulation outcome revealed that the use of higher SF value for the network having lots of nodes and it lowered the time consumption during the data collection process.

In 2020, Liao et al. [20] have suggested a model-driven deep reinforcement learning-based resource distribution system. Initially, the researchers developed a "Deep Neural Network (DNN)-based optimization network". This network was made up of "Alternating Direction Method of Multipliers (ADMM) iterative techniques". The CSI was assumed as a weight value using the above techniques. Here, the DNN-based optimization network was trained by the developed "Channel Information Absent Q-learning (CIAQ)" resource allocation algorithm. This algorithm consumed the minimum number of labeling data for completing the training process. Moreover, this algorithm controlled the discounting factor to optimize the spectral efficiency,fairness, and energy efficiency. Finally, the experimental assessment was carried out regarding energy efficiency, spectral efficiency, and fairness to identify the effectiveness of the offered system. In addition to that, the convergence speed of the proposed system was lower than the other techniques.

In 2020, Zhang et al. [21] offered a two-stage virtual network embedding algorithm with "Deep Reinforcement Learning (TS-DRL-VNE)" to provide the local optimal solution during the resource allocation process. The old Virtual Network Embedding (VNE) algorithm neglected the substrate network representation and the training mode in the resource allocation process. So it could be defeated by the "Full Attribute Matrix (FAM-DRL-VNE)" system. The old VNE algorithm could not able to maintain the underlying resources in the virtual network, and it was overcome by the "DRL-VNE algorithm aided on Matrix Perturbation Theory (MPT-DRL-VNE) system". The analytical comparison proved that the performance of the offered system was enhanced than the traditional techniques.

In 2019, Zhao et al. [22] have tendered reinforcement learning techniques. This technique was used to confirm the quality of service in the given cellular network. Here, the distributed optimization method depended on the multi-agent RL and the multi-agent DRL technique developed to defeat the computationally expensive problem in the huge action space. Afterward, the Dueling Double Deep Q-Network (D3QN) protocol was introduced to achieve the optimal policy by considering the attributes like reward function, state and action.The distributed user equipment achieved the global state space in the message-passing process. The experimental outcome suggested that the performance of the D3QN system was better than the traditional systems.

In 2023, Jouhari et al. [23] presented a Deep reinforcement learning system for resource distribution process. This system was utilized to maximize the energy efficiency in the LoRa structure. This structure was composed of flying GW and LoRa end devices so the network's lifetime could be greatly increased. The presence of SF and the wireless link was considered to allocate the SF to the wireless LoRa network. Here, the online resource allocation and the alteration of optimal policy were done with the aid of the flying GW system. Moreover, the action space was lowered to perform the efficient resource allocation. A comparative analysis was carried out to determine the efficiency of the system's efficiency, and it was proved that the energy efficiency of the developed system was better than the existing techniques.

In 2023, Minhaj et al. [24] introduced two independent learning techniques to allocate resources to the WAN. In this work, the combination of centralized and decentralized approaches was considered for allocating SF and transmission energy to the devices. The SF allocation process defeated the contextual bandit issue in the particular device. Moreover, the supervised learning issues were solved by transmission energy provided to the devices. Finally, the developed system was correlated with several existing approaches for evaluating the performance of the proposed model.

In 2022, Ashawa et al. [25] have explored a Long-Short Term Memory (LSTM) algorithm to carry out a better resource allocation process. The LSTM model was trained to get a better resource allocation outcome. The trained LSTM model was merged with the dynamic algorithm to solve the traffic issues in the cloud platform. The efficiency of the Monte Carlo Tree Search (MCTS) and the LSTM system were determined. The resource allocation process was carried out by comparing it with other load-balancing techniques. The analysis outcome showed that the accuracy of the proposed system was better than that of classical techniques. Additionally, the developed system consumes less consumes less time allocating resources to the WAN.

B. PROBLEM STATEMENT

The traditional methods used for allocating resources in WAN depend on the circumstance of the network as well. Therefore, efficient resource allocation, considering the dynamic nature of the network, is difficult to achieve. So, efficient resource allocation uses an advanced machine learning technique called DRL. The features and challenges of the existing reinforcement learning-based resource allocation method in WAN are given in Table 1. DQN [18] has enhanced the rate of consumption of energy in the LoRa network. However, the throughput offered by this method could be better. The time required to train this method is more. Also, there needs to be more data to pre-train this method. VNS and Minimum-cost spanning tree algorithm [19] method's operational and installation cost is lower. But, This method's performance is influenced by the presence of graph variants. DNN and CIAQ [20] can allocate the resources in future ultra dense network (UDN) more efficiently even with a restricted amount of CSI. The convergence capability of this method is faster. Yet, the training time required by this method is slightly higher. MPT-DRL-VNE, TS-DRL-VNE and FAM-DRL-VNE [21] Show better performance for allocating resources in VEN than other existing algorithms. The income-to-consumption ratio attained by this method is also higher. However, this method's reliability and efficiency to changes in the network are slightly below the requirement; hence, the agent's training could be more reliable. D3QN, MAQL, and MDP [22] attain higher generalization ability than existing DRL methods. This method achieved a faster convergence rate. Yet, this method needs more memory requirements. DNN and Proximal Policy Optimization (PPO) [23] obtained better energy efficiency in the LoRa network. Yet, this method's simultaneous operation of optimal resource allocation and data collection is impossible. Supervised machine learning and reinforcement learning [24] has enhanced the energy efficiency, good output, and the Packet Reception Rate (PRR) of dense LoRa networks. The time required by this method is lower. However, This method requires a feedback system for effective operation; hence, there is a possibility of uplink and downlink interference. LSTM and MCTS [25] provide the best resource allocation results when the traffic in the system is maintained stable. However, the generalization of the network is not achieved by this method. The energy consumption by the device connected by the individuals in the network is not computed by this method. So, an efficient resource allocation model by optimizing the network parameters is developed in this work using reinforcement learning techniques.

III. LOW-POWER WIDE AREA NETWORK MODEL AND PERFORMANCE CHALLENGES

A. NETWORK MODEL OF LORA IN LPWAN

The term LoRA denotes a long-range wireless communication system. The LoRA comprises network servers, gateways, unlicensed spectrum, and terminal nodes.The star-shaped communication network is formed by communication between the gateways and hop-1. The information is passed to the LoRA connection with the help of the gateway and the terminal components. The received data from the gateway and the terminal components is transferred via the gateway and the network server. The LoRA system has two layers: LoRAWAN and LoRA. The LoRA consists of a unique modulation layer that can easily transmit the data to long distances by consuming minimum power.

Bandwidth and the symbol duration are considered the most important terms in the LoRA network. The symbol duration is determined by Eq.1.

$$U_t = \frac{2^{TG}}{CX} \tag{1}$$

Here, the spreading factor is represented as TG and the bandwidth is denoted as CX, the symbol time duration is indicated as U_t.

The symbol duration is doubled based on the spreading factor and fixed bandwidth. The chance of collision between the packets mainly occurred due to the greater spreading factor value. In the signal encoding process, greater energy is consumed because of the rise of symbol duration. The message transmission time is increased with an increase in symbol duration. Therefore, communication can be easily done for a long distance with minimum shearing factor. The network model of LPWAN is indicated in Fig. 1.

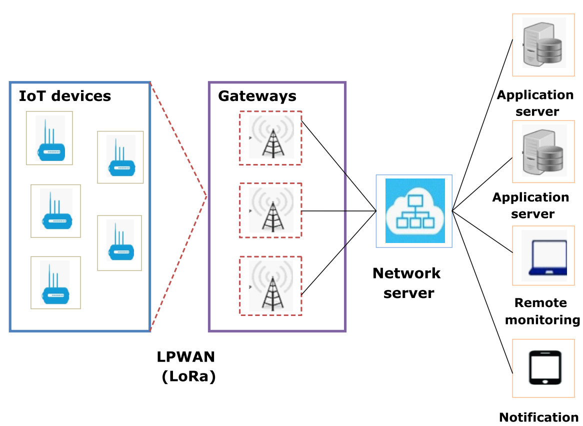


FIGURE 1. Network model of LPWAN.

TABLE 1. Authors, Methodology, Features, and Challenges.

Author [citation]	Methodology	Features	Challenges
Park *et al.* [18]	DQN	The energy consumption by the LoRa network is enhanced by this method.	• The throughput offered by this method is not good. • The time required to train this method is more. • There is no sufficient data to pre-train this method.
Araujo *et al.* [19]	VNS and Minimum-cost spanning tree algorithm.	The operational and installation cost of this method is lower.	This method's performance is influenced by the presence of graph variants.
Liao *et al.* [20]	DNN and CIAQ .	Even with a restricted amount of CSI, this method is able to allocate the resource in future UDN more efficiently. The convergence capability of this method is faster.	The training time required by this method is slightly higher.
Zhang *et al.* [21]	MPT-DRL-VNE,TS-DRL-VNE and FAM-DRL-VNE.	The performance achieved by this method for allocating resources in VEN is better than other existing algorithms. The income-to-consumption ratio attained by this method is also higher.	The reliability and the efficiency of this method to changes in the network are slightly below the requirement; hence the training of the agent is not much reliable.
Zhao *et al.* [22]	D3QN, MAQL and MDP.	The generalization ability of this method is higher than traditional deep reinforcement learning methods. This method achieved a faster convergence rate.	This method does not have enough memory requirements.
Jouhari *et al.* [23]	DNN and PPO.	This method obtained better energy efficiency in the LoRa network.	The simultaneous operation of optimal resource allocation and data collection is not possible in this method.
Minhaj *et al.* [24]	Supervised machine learning and reinforcement learning.	This method has enhanced the energy efficiency, goodput, and the PRR of dense LoRa networks. The time required by this method is lower.	This method requires a feedback system for effective operation; hence there is the possibility of uplink and downlink interference.
Ashawa *et al.* [25]	LSTM and MCTS.	This method provides the best resource allocation results when the traffic in the system is maintained stable.	• The generalization of the network is not achieved by this method. • The consumption of energy by the device connected by the individuals in the network is not computed by this method.

B. PROBLEMS RAISED IN THE LORA MODEL

Because of the uncontaminated Additive Links On-line Hawaii Area (ALOHA) and special modulation system, the LoRA can have the ability to provide long-range wireless communication with low power. The recognition of packets transferred from the gateway is very tough if the Received Signal Strength Indicator (RSSI) value of the LoRA gateway is higher. This issue is overcome by the adoption of an Adaptive Data Rate (ADR). ADR can manage energy consumption, data rate, and symbol duration, and it can result is the prevention of negative effects in the LoRA terminal. Initially, 20 data packets are gathered by the ADR system, and further, the data packets are examined for the bracing process, and it can be achieved by increasing the spreading factor and lowering the transmission power. As a result of this scenario, the strength of the transmission signal is increased, and the chance of packet collision is greatly lowered.

C. DESCRIPTION OF THE DEVELOPED MODEL AND ITS MOTIVATION

The well-suited transmission power and spreading factor must be selected for minimizing the transmission power in the LoRA network. The appropriate spreading factor values can enhance the throughput of the LoRA. Nevertheless, the loss rate is greatly increased due to the attributes selected by the ADR. The transmission power and the enciphering system are vary based on the distance in the middle of LoRA nodes and its gateway. In the LoRAWAN system, the network resource issue is solved by ADR and other techniques. Here, a very small amount of samples and the user-defined threshold value are considered for identifying the attribute values. Therefore, the complexity of the algorithm is greatly lowered. These attributes are utilized to enhance the total throughput value of the network. However, energy consumption is greatly increased in the resource allocation process because of the developed attributes.To solve these problems, we developed the reinforcement-based optimal resource allocation system. The pictorial representation of the reinforcement learning-based optimal resource allocation system is shown in Fig. 2.

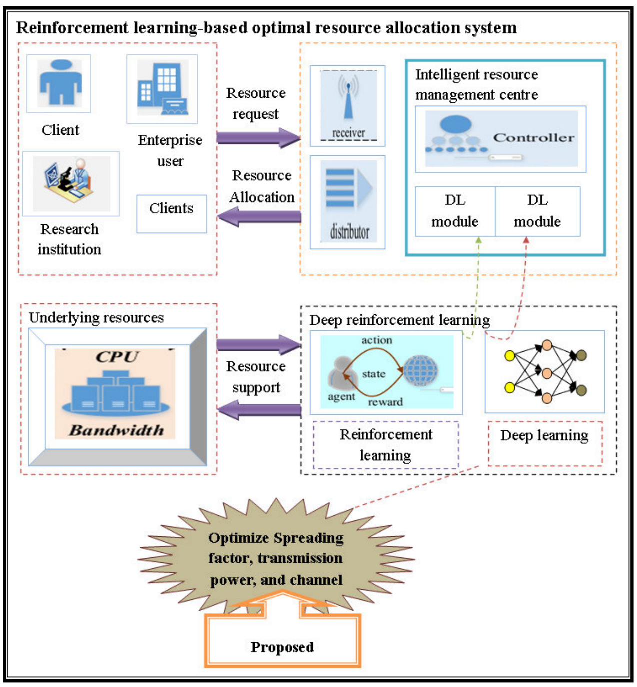


FIGURE 2. Pictorial representation of reinforcement learning-based optimal resource allocation scheme in low power WAN.

The reinforcement learning-based optimal resource allocation system is developed for optimally allocating the resources to the LPWAN with minimum power consumption. The developed system is also used to enhance the enlarged packet arrival rate. The central network server receives the data about the LoRA terminal. The attributes in the network servers are managed by the reinforcement learning technique for the optimal allocation of resources to the LPWAN. The LoRA is a star-shaped structure where the centralized reinforcement learning system and the multi-agent reinforcement learning system are combined to determine the spreading factor, channel and transmission power for the resource allocation process. The information about each terminal is identified by the controller of the centralized reinforcement learning system, and it is used to allocate the resources to a terminal of the LPWAN. The total number of terminals is considered in the resource allocation process. The action space depends upon the total count of terminals in the LPWAN. Here, the controller is developed with the aid of a reinforcement learning agent. The network terminal consists of several LoRA, which can be elaborately learned by the reinforcement agents created by the developed model. In the developed optimal resource allocation model, the reinforcement learning technique is considered to compute the attributes like spreading factor, channel and transmission power.Consequently, the offered HC-EVOA is adopted to select the spreading factor, channel and transmission power optimally. Based on the optimally selected attributes, the resources are allocated to the nodes of the LPWAN. The experimental analysis is conducted to examine the performance of the developed model.

IV. REINFORCEMENT LEARNING PARAMETER OPTIMIZATION IN LPWAN WITH HC-EVOA

A. COATI OPTIMIZATION ALGORITHM (COA)

The COA [26] replicates the characteristics of the coati animal. The COA algorithm is adopted to defeat real-world optimization problems.

Motivation and the characteristics of the coati animal: Coati animals are distributed in the United States region, and they have noses in the upward direction, small ears, black paws and tails. This animal passes signals by using its tail. The total length of the adult coati animal is in the range of 33 to 69 cm, and they weigh about 2 to 8 kilograms. The male coati animal can mature twice the size of female coati animals. The coati animal can take the food in the form of small lizards, rats and eggs of the crocodiles. But, it takes lizards as the most wanted food. The lizards spend most of their lifetime on trees. So, the coati animals forge the lizard by forming a team. In this consequence, one coati animal mounts on the top of the tree to reach near to the lizard, and the coati animal tries to fall down the lizard on the ground portion. In this situation, other team members of the coati animal started their attacking process. Eagles, dogs, other animals, etc, may also hunt the coati animals.

In COA, the population's individuals are considered the coati members. The judgment variable is determined from the location of the coati members in the search space. Therefore, Coatti's position is the candidate solution to the particular issues. The initialization of the coati animals in the search space is based on the location, and expressed in Eq.2.

$$Y_a : y_{a,b} = lbound_b + s(ubound_b - lbound_b), \quad here, a = 1, 2..O \quad b = 1, 2..n \tag{2}$$

Here, the location of the coati animal at a^{th} position is represented as Y_i, the decision variable is indicated as n, and the count of coati animals is indicated as O, the judgment variable of b is indicated as $Y_{a,b}$, the upper and the lower bounds are indicated as $ubound_b$, $lbound_b$ and the arbitrary real number is represented as s in [0, 1].

Arithmetic design of COA: The coati animals have two main characteristics attacking characteristics and escaping characteristics, and they are detailed in upcoming sections.

Attacking characteristics: This attacking characteristic is initially used to upgrade the position of the population of coati animals. Here, the coati animal started mountaineering on the top of the tree to panic the lizard, and they fell to the ground portion. Here, the remaining team members of the coati animal forage the lizard fell down on the ground portion. In this scenario, the coati must be tried to change the position on the search space.

Here, the location of the lizard is considered the location of the best coati animal. In the hunting process of coati animals, the total team of coati animals is divided into two parts: half of the team is located on the ground portion, and the remaining team members of coati animals mount on the tree to panic the lizard. Here, the location of the coati animals mounted on the tree is identified by Eq. 3.

$$Y_a^{q^1} : y_{a,b}^{q^1} = y_{a,b} + s(lizard_b - J.y_{a,b}), \quad here, a = 1, 2..\frac{O}{2} \quad b = 1, 2..n \tag{3}$$

The lizards are located arbitrarily after the fell-down process. The coati animals started their displacement process based on the lizard animals' arbitrary position, formulated using Eq. 4, Eq. 5 and Eq. 6.

$$lizard^H : lizard_k^H = lbound_k + s(ubound_b - lbound_b), \quad b = 1, 2..n \tag{4}$$

$$Y_a^{q^1} : y_{a,b}^{q^1} = \begin{cases} y_{a,b} + s(lizard_b^H - J.y_{a,b}), G_{lizard^H} < G_a \\ y_{a,b} + s(lizard_b^H) - J.y_{a,b}), else \end{cases} \tag{5}$$

$$for\ a = \left[\frac{O}{2}\right] + 1, \left[\frac{O}{2}\right] + 2, \ldots O, and\ b = 1, 2..n \tag{6}$$

If the new position enhances the objective function, then the new position is upgraded. If the objective function is not enhanced, then the location of the coati animals not updated. The condition a = 1,2..O is determined by Eq. 7.

$$Y_a = \begin{cases} Y_a^{q^1} + G_a^{q^1} < G_a \\ Y_a \text{ else} \end{cases} \tag{7}$$

Here, the coati's updated position is indicated as $Y_a^{q^1}$, the arbitrary original number is represented as s in [0,1], the position of lizards is indicated as a $lizard$, the location and the objective function of the lizard on the ground portion is signified as $lizard^H$, and $G_a^{q^2}$ respectively, the integer value is indicated as J in [1,2], and the floor function is represented as $\lfloor \bullet \rfloor$.

Escaping behaviour: The coati animal escapes from its enemies, and this behaviour is used to update the location of the coati animal. The coati escapes from animals like dogs and eagles for safety purposes, and the new position of the coati animal is closer to the present location.

The coatis are located in a random position created near the present location of the coati animal. The random position is created by Eq. 8 and Eq. 9.

$$lbound_b^{loc} = \frac{lbound_b}{u}, \ ubound_b^{loc} = \frac{ubound_b}{u} \quad here, u = 1, 2..U \tag{8}$$

$$Y_a^{q^2} : y_{a,b}^{q^2} = y_{a,b} + (1 - 2s)(lbound_b^{loc}) + s(ubound_b^{loc} - lbound_b^{loc}), Where, \quad a = 1, 2..O, b = 1, 2..n \tag{9}$$

Here, the objective position is enhanced by the new position, then only the new position is accepted and it is stimulated in Eq. 10.

$$Y_a = \begin{cases} Y_a^{q^2} + G_a^{q^2} < G_a \\ Y_a \text{ else} \end{cases} \tag{10}$$

Here, the local upper and the lower bound is represented as $ubound_b^{loc}$, $lbound_b^{loc}$, the innovative location of the coati animal is indicated as $Y_a^{q^2}$, at b^{th} dimension, the location n of the coati animal is indicated as $Y_{a,b}^{q^2}$, the objective function is indicated as $G_a^{q^2}$, the iteration counter is represented as U, and the random number is represented as S in [0,1].

The pseudocode of the COA is indicated in Algorithm.1.

B. ENERGY VALLEY OPTIMIZER(EVO)

This algorithm is designed with advanced physics principles. This algorithm gives superior solutions to huge and tiny optimization issues.

Arithmetical form: In EVO [27] algorithm, the particles are considered as the candidate solution, and it is indicated as Y_a and it is presented in Eq. 11.

$$Y_a^b = y_{a,min}^b + s(y_{a,max}^b - y_{a,min}^b) \begin{cases} a = 1, 2..o \\ b = 1, 2..e \end{cases} \tag{11}$$

Here, the b^{th} decision variable is represented as y_a^b, and it is used to find the first location of the a^{th} candidate, the size of

Algorithm 1 COA

Require: Input the optimization data.

Initialize maximum count of iteration and total count of coati animals
for each iteration **do**
  Upgrade the position of the lizard.
  Exploration phase (1st phase)
  for each coati animal **do**
    Identify the location of coati animal.
    Upgrade the position of coati.
  end for
  Exploitation phase (2nd phase)
  Determine the local bound value for all variables.
  Update the new location of the coati.
end for
Achieve best solution.

the issue is signified as e for a^{th} the candidate, the upper and the lower bound of the variable is indicated as $y^b_{a,max}$, $y^b_{a,min}$, and the uniformly dispersed arbitrary number is indicated as s *in* [0,1].

The enrichment bound of all particles is determined in the second phase of the algorithm. The enrichment bound determines the difference between the poorest and the best neutron. To calculate the enrichment bound value, one must identify the objective value for all particles, available in the form of neutron enrichment levels and mathematically expressed in Eq. 12.

$$FC = \frac{\sum_{a=1}^{o} OFM_a}{o} \quad a = 1, 2..o \tag{12}$$

Here, the enrichment bound is represented as FC, and the a^{th} particle neuron enrichment level is represented as OFM_a.

The objection function identification is utilized to identify the stability level of the particles, and it is specified in Eq. 13.

$$TM_a = \frac{OFM_a - CT}{XT - CT} \quad a = 1, 2..o \tag{13}$$

Here, the best and the worst stability levels of the particles are represented as CT and XT, respectively, the a^{th} particles stability level is represented as TM_a.

If the neutron enrichment level of the particle is greater than the enrichment bound in the loop process of EVO, then the neutron-to-proton ratio of the particle is too large; therefore, the alpha and the beta parameters are considered in the decay process. The stability bound value is greater than the stability of the particle, and then the gamma and alpha decay processes are considered for maintaining the stability of the particle. During the physical reaction process, the stability of the particle is maintained with the assistance of alpha rays. The new positions of the particles are identified based on the above scenario, and here, two numbers of alpha indexes are created, and they are in the interval of [1,e] *and* [1,*indexalpha*] respectively. Here, the decision variables of the solution candidates are considered as the emitted rays. The emissions of rays are controlled with the aid of the best stability level of the particles which is given in Eq. 14.

$$Y_a^{New1} = Y_a(Y_{CT}(y_a^b)), \begin{cases} a = 1, 2..o \\ b = \text{indexalpha2} \end{cases} \tag{14}$$

Here, the present position vector for the a^{th} particle is indicated as Y_a, the innovative particle in the cosmos, is represented as $Y_a^N ew1$ the emitted ray at the position b^{th} is signified as y_a^b, and the location vector of the particle having best stability value is indicated as Y_{CT}.

Nevertheless, the gamma decay process also enhances the particle's stability level. The new candidate solution is developed for the position upgrading process. In this process two integers, gamma index 1 and gamma index two, are developed and stuck in the range of [1,e] *and* [1,*indexgamma*] respectively. Both of the indexes are used to identify the photons of the particle. The decision variables of the candidate solution are considered as the photons of the particle. The neighbouring particles can have the ability to remove the photons of the particles. The distance in the middle of two particles is calculated by Eq. 15.

$$E_a^c = \sqrt{(y_2 - y_1)^2 + (z_2 - z_1)^2}, \quad \begin{cases} a = 1, 2 \ldots o \\ c = 1, 2 \ldots o - 1 \end{cases} \tag{15}$$

Here, the coordinates of the particles are indicated as $(y_2 - y_1)$,$(z_2 - z_1)$, the distance in the middle of a^{th} particles and the c^{th} adjacent particles are represented as E_a^c.

The subsequent solution candidate is developed for updating the position, and it is given in Eq. 16.

$$Y_a^{New2} = Y_a(Y_{oh}(y_a^b)), \quad \begin{cases} a = 1, 2..o \\ b = \text{index gama 2} \end{cases} \tag{16}$$

Here, the present position vector for the a particle is indicated as Y_a, the innovated particle in cosmos is represented as Y_a^{New2} the position vector of the adjacent particle is indicated as Y_{oh} and the decision variable at b^{th} position is indicated as y_a^b.

The beta decay process is considered if the stability bound value is greater than the stability of the particle. The stability of the particle is enhanced by removing the beta rays from the particles. Further, the location upgrading process is done by considering the best stability level and the centre of the particles, which is depicted in Eq. 17 and Eq. 18.

$$Y_{DQ} = \frac{\sum_{a=1}^{o} Y_a}{o}, \quad a = 1, 2..o \tag{17}$$

$$Y_a^{New1} = Y_a + \frac{(s_1 \bullet Y_{CT} - S_2 \bullet Y_{DQ})}{TM_a} \quad a = 1, 2..o \tag{18}$$

Here, the position vector of the particle has the best stability level and it is indicated as Y_{CT}, the position vector of the adjacent particle is represented as Y_{DQ}, the b^{th} particles existing and the forthcoming position is indicated as Y_a, *and* Y_a^{New1} respectively, a^{th} particles stability level is represented asTM_a, the random numbers are indicated as s_1, *and* s_2 respectively and they are available in the interval of [0,1].

The position upgrading process is done to improve the exploitation and exploration phases of the EVO algorithm. This position updating process is stimulated in Eq. 19.

$$Y_a^{New2} = Y_a + (s_3 \times Y_{CT} - s_4 \times Y_{oh}) \;\; a = 1, 2..o \tag{19}$$

Here, the random numbers are represented as s_3 *and* s_4 in [0,1], the present and the forthcoming position vector of the a^{th} particles indicated as Y_a, and Y_a^{New2} respectively.

The neutron-to-proton ratio of the particle is taken as small if the enrichment bound of the particles is greater than the neutron enrichment level of the particle. So, this process sent the particle to perform the positron emission process. Consequently, it causes random displacement in the particles, as formulated in Eq. 20.

$$Y_a^{New} = Y_a + s, \;\; a = 1, 2..o \tag{20}$$

Here, at the given search space, the a^{th} particles forthcoming and the existing position vector is represented as Y_a^{New}, Y_a, respectively.

The pseudocode of the EVO algorithm is indicated in Algorithm.2.

Algorithm 2 EVO Algorithm

Find the location of all candidate solutions present in the search space.
Determine the fitness of the solution candidates in terms of *OFMa*.
while count of function evaluation < function evaluation **do**
  Find the enrichment bound *FC* of all particles.
  Find the particle based on the best stability bound.
  for $j = 1$ to o **do**
    Determine the stability bound of the particle.
    Find the neutron enrichment level of the particle *OFMa*.
    if *OFM* < *FC* **then**
      Find the stability bound of the particle.
    end if
    if *TMa* > *TC* **then**
      Find index alpha I and index alpha II.
    end if
  end for
end while
Achieve best solution.

C. IMPLEMENTED HC-EVOA

The HC-EVOA algorithm, which has been developed, consists of two traditional algorithms: EVO and COA. It is employed within the HC-EVOA-based optimal resource allocation system to select the optimal spreading ratio, transmission power, and channels for allocating resources to the terminal nodes of the LPWAN. This algorithm aims to minimize SINR during the resource allocation process.

The EVO is a physics-based algorithm for addressing complex real-world optimization issues. Still, it may encounter challenges in local optimization. On the other hand, the COA algorithm, which is bio-inspired, strikes a balance between exploitation and exploration phases to provide optimal solutions but may not address real-time issues.

The developed HC-EVOA algorithm resolves these issues by incorporating the initial positions obtained from the EVO and COA algorithms. It is mathematically represented in Eq. 21.

$$np = olpo + std(Y_a^{q^2} + Y_a^{New}) \tag{21}$$

Here, the term *olpo* indicates the old position. The position obtained from the COA and EVO algorithm is represented as $Y_a^{q^2}$ and Y_a^{New}, respectively. The updated position achieved by the developed HC-EVOA is indicated as *np*. Becuase of this algorithm, the SINR is reduced in the resource allocation process.

The schema chart and the pseudocode of the developed HC-EVOA are denoted in Fig. 3 and Algorithm.3.

Algorithm 3 Developed HC-EVOA

Initialize the number of population *popnum* and the maximum number of iterations *mter*.
Update the position of the lizards.
Find the fitness value of the solution candidates.
if *OFM* < *FC* **then**
  Evaluate the stability bound of the particle.
  for $t = 1$ to *mter* **do**
    for $i = t$ to *popnum* **do**
      Calculate the position $Y_a^{(q^2)}$ by Eq. (9).
      Upgrade the position by adopting COA.
      Calculate the position Y_a^{New} by Eq. (20).
      Upgrade the position by adopting EVO.
      Calculate $np = olpo + \mathrm{std}(Y_a^{(q^2)} + Y_a^{New})$.
    end for
  end for
end if
Obtain the best solution.

V. EFFICIENT OPTIMAL RESOURCE ALLOCATION USING HYBRID MODEL

A. DEEP REINFORCEMENT LEARNING

To achieve the best results in various situations, reinforcement learning is utilized. The environment's condition is also determined using the deep reinforcement learning technique. The Markov Decision Process (MDP) is developed from the reinforcement learning technique. The MDP's condition is determined by the agents, and it is also adopted to enhance future rewards over time. The total future reward is expressed in Eq. 22.

$$H_U = \sum_{J=u}^{U} S_{j+1} \tag{22}$$

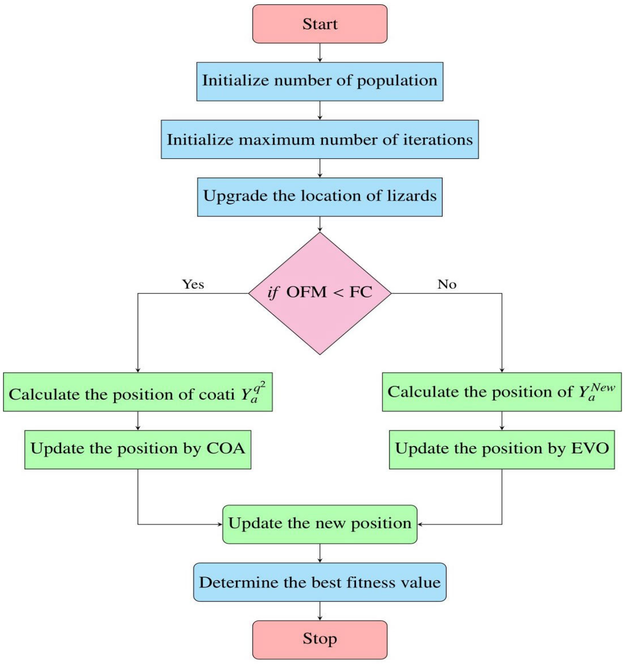


FIGURE 3. Schema chart of the developed HC-EVOA.

The particular action b_u is executed to classify the state t_u at the time u, and also the value s_{u+1} is also generated with the aid of the action. As a result of this process, the innovated state is making communication with the tuple and it is represented as $t_u, b_u, s_{u+1}, t_{u+1}$.These tuples are used by the agent for the learning process. The activation-valued function is taken into consideration during the learning process, and it is given in Eq. 23.

$$R(t, b) = F^{\pi}(H_u|T_u = t,\ B_u = b) \tag{23}$$

Here, the action state is used for getting the predicted value. Based on the agent's performance in examining the current policy, the current policy is divided into two types such as on policy and off-policy. With the help of off-policy reinforcement learning, better optimal values are elected. The off-policy reinforcement learning is also called as Q learning and it tries to learn the policy from the greed policy system. The upgraded function value in the Q learning process is depicted in Eq. 24.

$$R(t_u, b_u) = R(t_u, b_u) + A(s_u + \lambda_{\max_{b'}} R(t'_u, b'_u) - R(t_u, b_u)) \tag{24}$$

The loss function is explained by connecting the Deep Q network (DQN) with the attributes of the artificial neural network. To lower the cost of the loss function, this process is repeated for a time, and it is presented in Eq. 25 and Eq. 26.

$$M_j(\theta_j) = F_t, b, s, t[z_j^{DQN} - R(t, b : \theta)^2] \tag{25}$$

$$z_j^{DQN} = s + A\ max_b, (t', b'; \theta_j^-) \tag{26}$$

The DQN resolves the huge size space problem accompanied by the deep learning technique.

B. REINFORCEMENT LEARNING AGENT

The reinforcement work consists of several counts of LoRa nodes and only one LoRa gateway. Both of the components are stead in the 1500 × 1500 size grid. The LoRa gateway is placed in the middle of the grid, and the remaining nodes are disseminated on the same grid. Based on the ADR, the gateway and the nodes alter the transmission power and the spreading factor. For all the nodes of the agent, the LoRa gateway is assumed as a container, and this assumption is made for achieving the transmission power and the spreading factor via the reinforcement learning technique. Here, the agents are utilized, more nodes are present in the network, and it is grouped together so the agent can easily control it. Finally, the optimal allocation of resources highly balances the energy harvesting and energy consumption at sensors. It highly stabilizes the wide area network by maximizing the user's utility. The diagrammatic illustration of the reinforcement learning agent is represented in Fig. 4.

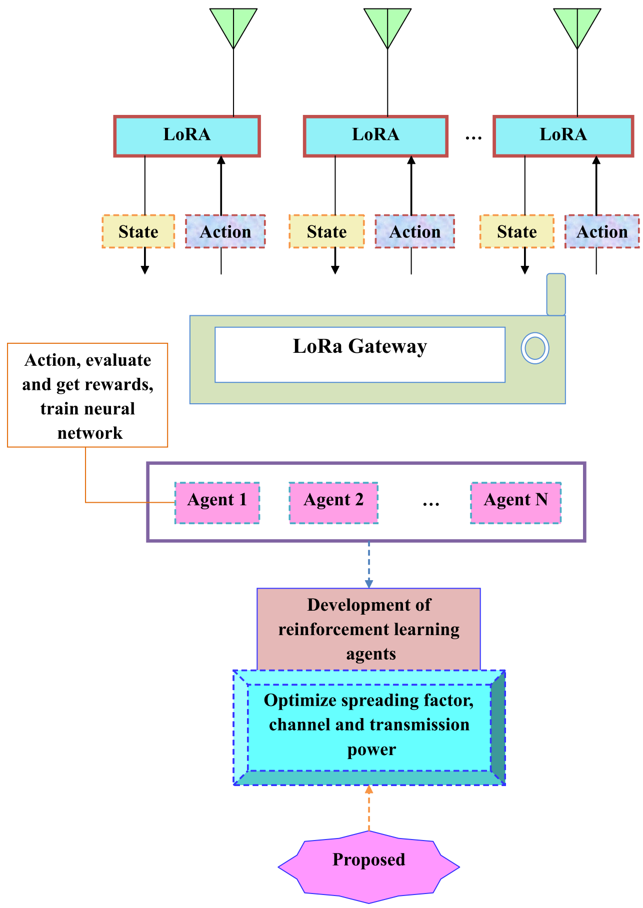


FIGURE 4. Diagrammatic illustration of the reinforcement learning agent.

The reinforcement learning agent is used to analyze the LoRA network, and it identifies the attributes of the network, and here, the optimization is easily done by the Markov Decision Process (MDP). At the time zone u, the node data is indicated as t_u, the terminal nodes present in the network are represented as O, from any one of the terminal nodes, the elected attributes are represented as b_u, and the compensation value achieved from the network attributes is indicated as s_u. The state of the LoRA is defined using the above attributes. The details of reinforcement learning are studied with the aid of tuples t_u, b_u, s_{u+1}, t_{u+1}. The DQN is used to form the reinforcement agents of the network.

C. THE OBJECTIVE FUNCTION OF THE DEVELOPED MODEL

The objective function of the HC-EVOA-based optimal resource allocation system is used to lower the SINR value during the resource allocation process, and it is expressed in Eq. 27.

$$F = \underset{\left\{c_n^{\text{LPWAN}}, s_F^{\text{LPWAN}}, T_p^{\text{LPWAN}}\right\}}{\arg\min} (\text{SINR}) \tag{27}$$

Here, the term c_n^{LPWAN} is represented as the channel, and it lies in the interval of [1 *to number of channel*], spreading factor and the transmission power is indicated as s_F^{LPWAN}, T_P^{LPWAN}, and they lies in the middle of [0.01, 0.99] *and* [2, 128], respectively. The value of SINR is identified by Eq. 28.

$$SINR = \frac{||I_{JJ} X_{JJ}||^2}{\sum_{L=1, L\neq J}^{o} ||I_{JJ} X_{JJ}||^2 + K^2} \tag{28}$$

Here, the channel vector of the user is represented as I_{JJ}, the unit norm beam factor is indicated as X_{JJ}, and the SNR value per user is represented as $\frac{1}{K^2}$.

VI. RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

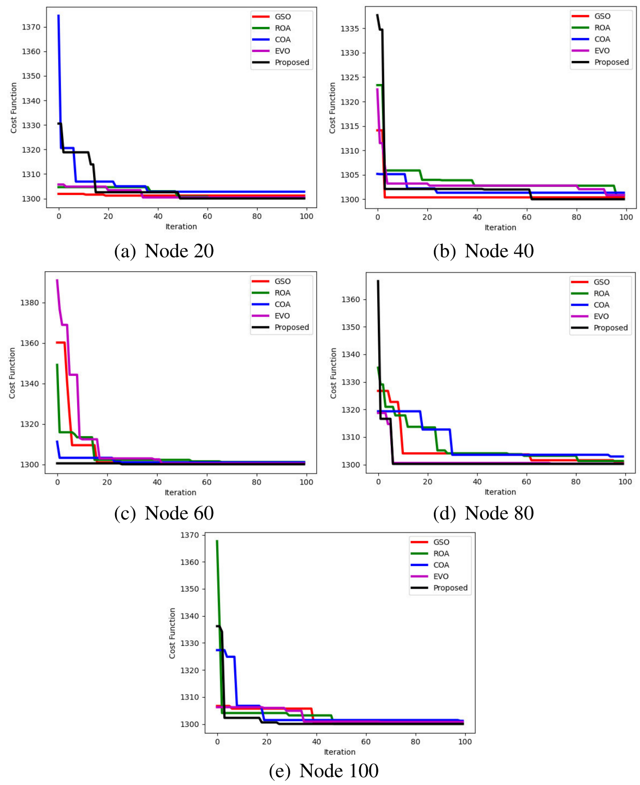
The proposed reinforcement learning-based optimal resource allocation system was executed in the Python platform. In this analysis, the count of the population was adopted as 10 and the remaining attributes, such as maximum iteration and length of the chromosome were adopted as 50 and 3, respectively, Conventional heuristic algorithms like "Galactic Swarm Optimization(GSO) [28], Rain Optimization Algorithm (ROA) [29], COA [26] and EVO [27]" were considered for finding the efficacy of the developed model.

B. CONVERGENCE ASSESSMENT

Fig. 5 stimulates the convergence examination of the HC-EVOA-based optimal resource allocation system. At the node value of 20, the cost function of the HC-EVOA-based optimal resource allocation system is 0.07% lower than GSO, 0.15% lower than ROA, 0.23% lower than COA and 0.30% lower than EVO at iteration 80. So, it is proved that the cost requirement of optimal resource algorithm based on reinforcement learning is more efficient than the meta-heuristic algorithm-based optimization.

C. EFFICIENCY ASSESSMENT BY VARYING THE NUMBER OF NODES

The efficiency evaluation of the invented HC-EVOA-based optimal resource allocation system using the number of epochs is presented in Fig. 6. Here, the analysis takes the number of nodes to produce accurate results. At the number of nodes is 80, the throughput value of the HC-EVOA-based optimal resources allocation system is improved by GSO, ROA, COA and EVO with 18.18%, 10.16%, 54.76% and 6.55%. Therefore, the comparative outcome showed that the throughput of the developed HC-EVOA-based optimal resource allocation system is enhanced than the old algorithms.



(a) Node 20 (b) Node 40 (c) Node 60 (d) Node 80 (e) Node 100

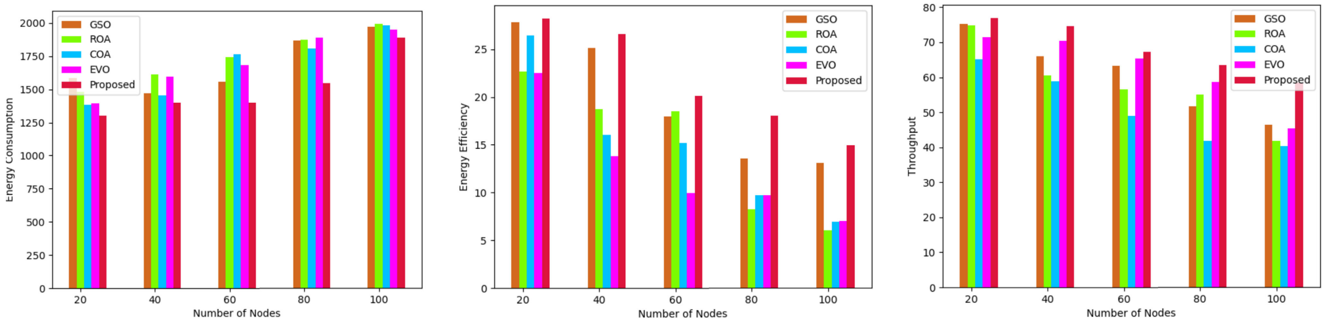
FIGURE 5. Cost function assessment of the developed Reinforcement Learning-based optimal resource allocation system for (a) Nodes 20, (b) Nodes 40, (c) Nodes 60, (d) Nodes 80, (e) Nodes 100.

D. ENERGY CONSUMPTION ANALYSIS

The energy consumption analysis of the HC-EVOA-based optimal resource allocation system is shown in Fig.7. The comparison is done by using two attributes like spreading factor and transmission power. At transmission power of 8, the developed HC-EVOA-based optimal resource allocation system obtains energy consumption as 12.86% lower than GSO, 19.29% lower than ROA, 2.76% lower than COA and 2.81% lower than EVO. Hence, the comparative results depicted that the energy consumption of the proposed HC-EVOA-based resource allocation system was greatly reduced compared to the other algorithms.

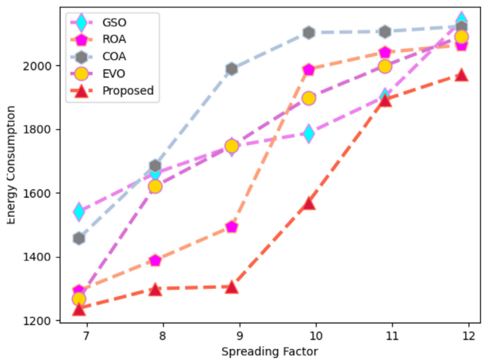
E. THROUGHPUT ANALYSIS

Fig. 8 gives the throughput analysis of the HC-EVOA-based optimal resource allocation system. The throughput value of

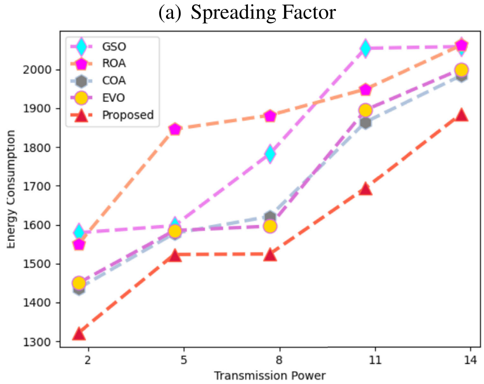


(a) Energy consumption (b) Energy efficiency (c) Throughput

FIGURE 6. **Efficiency assessment of the developed Reinforcement Learning-based optimal resource allocation using a number of nodes for (a) Energy consumption, (b) Energy efficiency, (c) Throughput.**

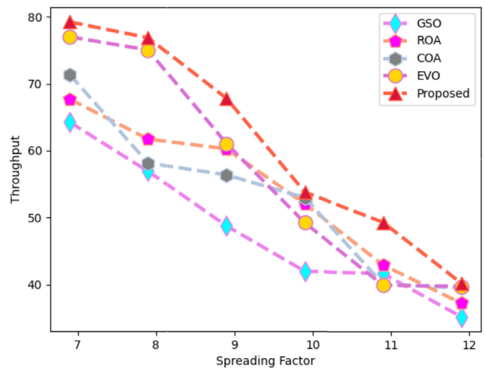


(a) Spreading Factor

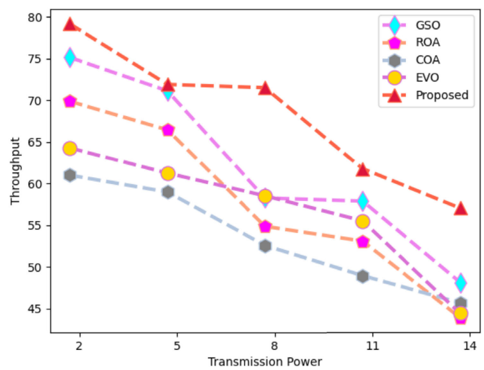


(b) Transmission Power

FIGURE 7. **Energy consumption analysis of the Reinforcement Learning-based optimal resource allocation process for (a) Spreading factor, (b) Transmission power.**



(a) Spreading Factor



(b) Transmission Power

FIGURE 8. **Energy consumption analysis of the Reinforcement Learning-based optimal resource allocation process for (a) Spreading factor, (b) Transmission power.**

the HC-EVOA-based optimal resources allocation system is greater than the GSO, ROA, COA and EVO with 23.80%, 15.55%, 26.82% and 30% at a spreading factor of 11. So it is confirmed that the throughput value of the proposed HC-EVOA-based optimal resource allocation system is enhanced then the classical algorithms.

F. STATISTICAL EXAMINATION

The statistical examination of the developed HC-EVOA-based on the resource allocation system is signified in Table: 2. At the number of nodes 100, the standard deviation value of the HC-EVOA-based optimal resource allocation system is 54.96% more than GSO, 26.65% more than ROA, 11.25% more than COA and 59.93% more than EVO in terms of standard deviation value. Therefore, statistical results proved that the performance of the developed HC-EVOA-based optimal the resource allocation system is superior to the ancient techniques.

The proposed model successfully addresses the critical challenge of resource allocation in LPWANs by combining reinforcement learning and HC-EVOA. Results consistently demonstrate its effectiveness, showcasing notable reductions

TABLE 2. Statistical examination of the developed reinforcement learning-based resource allocation system.

ALGORITHMS	GSO [28]	ROA [29]	COA [26]	EVO [27]	HC-EVOA
Node 20					
BEST	1301.206	1300.188	1302.784	1300.459	1300.038
WORST	1301.898	1304.648	1374.51	1305.742	1330.565
MEAN	1301.314	1302.222	1305.52	1301.779	1303.856
MEDIAN	1301.206	1300.771	1302.784	1300.459	1300.038
STD DEVIATION	0.231879	2.031826	8.146921	1.900516	7.001677
Node 40					
BEST	1300.383	1300.835	1301.334	1300.843	1300.008
WORST	1314.108	1323.358	1305.188	1322.483	1337.67
MEAN	1300.795	1304.024	1301.904	1303.038	1302.295
MEDIAN	1300.383	1302.776	1301.342	1302.791	1301.995
STD DEVIATION	2.341176	3.610535	1.230162	2.427789	5.963896
Node 60					
BEST	1300.936	1301.11	1300.97	1300.429	1300.029
WORST	1360.169	1349.151	1311.169	1390.893	1300.477
MEAN	1304.783	1303.935	1301.574	1307.471	1300.146
MEDIAN	1300.936	1302.192	1300.97	1300.429	1300.029
STD DEVIATION	12.40257	6.401766	1.334046	17.91578	0.196382
Node 80					
BEST	1300.733	1301.278	1302.934	1300.363	1300.241
WORST	1326.73	1335.157	1319.333	1318.731	1366.557
MEAN	1305.037	1306.691	1307.455	1301.524	1301.722
MEDIAN	1304.054	1304.088	1303.55	1300.581	1300.241
STD DEVIATION	6.527721	6.981518	6.301241	4.036261	7.427645
Node 100					
BEST	1300.118	1300.065	1301.172	1300.773	1300.027
WORST	1306.666	1367.621	1327.336	1306.213	1336.184
MEAN	1302.437	1302.756	1303.986	1302.601	1301.465
MEDIAN	1300.357	1300.065	1301.449	1300.906	1300.027
STD DEVIATION	2.729626	7.655692	6.728397	2.424382	6.040008

in energy consumption (up to 19.29%) and substantial improvements in throughput (up to 54.76%) compared to existing algorithms. These findings underscore the model's potential to significantly enhance the efficiency and performance of LPWANs, making it a valuable contribution to the field.

VII. CONCLUSION

The reinforcement learning-based optimal resource allocation system was developed to allocate the resources to the terminal nodes of the LPWAN. Here, reinforcement learning was adopted for computing the attributes like spreading factor, transmission power and channel for the resource allocation process. After that, the developed HC-EVOA algorithm was adopted for optimizing the spreading factor, transmission power and channel, and the proposed algorithm was utilized to lower the SINR ratio in the resource allocation process. Finally, the resources were allocated to the terminal nodes of the LPWAN based on the optimized attributes. At the number of nodes 80, the throughput value of the HC-EVOA-based optimal resources allocation system was improved with GSO, ROA, COA and EVO by 18.18%, 10.16%, 54.76% and 6.55%.These results validate the system's ability to enhance network performance, reduce energy consumption, and maintain stability under varying network conditions. It highlights our approach's practical importance and potential influence on LPWAN applications, establishing it as a valuable contribution to computer networking and resource optimization. Future works consider analyzing the energy model for selecting the antenna and its impacts on the range and reliability.

REFERENCES

[1] A. Nauman, M. A. Jamshed, Y. A. Qadri, R. Ali, and S. W. Kim, "Reliability optimization in narrowband device-to-device communication for 5G and beyond-5G networks," *IEEE Access*, vol. 9, pp. 157584–157596, 2021.

[2] Y. Cui, L. Du, H. Wang, D. Wu, and R. Wang, "Reinforcement learning for joint optimization of communication and computation in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 13062–13072, Dec. 2021.

[3] P. Almasan, S. Xiao, X. Cheng, X. Shi, P. Barlet-Ros, and A. Cabellos-Aparicio, "ENERO: Efficient real-time WAN routing optimization with deep reinforcement learning," *Comput. Netw.*, vol. 214, Sep. 2022, Art. no. 109166.

[4] G. Sun, G. O. Boateng, D. Ayepah-Mensah, G. Liu, and J. Wei, "Autonomous resource slicing for virtualized vehicular networks with D2D communications based on deep reinforcement learning," *IEEE Syst. J.*, vol. 14, no. 4, pp. 4694–4705, Dec. 2020.

[5] J. Bae, J. Lee, and S. Chong, "Learning to schedule network resources throughput and delay optimally using Q^+-learning," *IEEE/ACM Trans. Netw.*, vol. 29, no. 2, pp. 750–763, Apr. 2021.

[6] T. Xu, M. Zhao, X. Yao, and Y. Zhu, "An improved communication resource allocation strategy for wireless networks based on deep reinforcement learning," *Comput. Commun.*, vol. 188, pp. 90–98, Apr. 2022.

[7] G. Zhou, R. Wen, W. Tian, and R. Buyya, "Deep reinforcement learning-based algorithms selectors for the resource scheduling in hierarchical cloud computing," *J. Netw. Comput. Appl.*, vol. 208, Dec. 2022, Art. no. 103520.

[8] M. Mohammadi, S. A. Motamedi, and S. Sharifian, "Drops on surface optimization (DSO): A new reinforcement learning based metaheuristic algorithm for virtual network functions placement in distributed cloud architectures," *Comput. Netw.*, vol. 219, Dec. 2022, Art. no. 109459.

[9] Z. Chen, L. Zhang, X. Wang, and K. Wang, "Cloud–edge collaboration task scheduling in cloud manufacturing: An attention-based deep reinforcement learning approach," *Comput. Ind. Eng.*, vol. 177, Mar. 2023, Art. no. 109053.

[10] Y. Wang, Y. Li, T. Wang, and G. Liu, "Towards an energy-efficient data center network based on deep reinforcement learning," *Comput. Netw.*, vol. 210, Jun. 2022, Art. no. 108939.

[11] B. Sellami, A. Hakiri, S. B. Yahia, and P. Berthou, "Energy-aware task scheduling and offloading using deep reinforcement learning in SDN-enabled IoT network," *Comput. Netw.*, vol. 210, Jun. 2022, Art. no. 108957.

[12] C. Chen, Y. Zhang, Z. Wang, S. Wan, and Q. Pei, "Distributed computation offloading method based on deep reinforcement learning in ICV," *Appl. Soft Comput.*, vol. 103, May 2021, Art. no. 107108.

[13] T. Zheng, J. Wan, J. Zhang, and C. Jiang, "Deep reinforcement learning-based workload scheduling for edge computing," *J. Cloud Comput.*, vol. 11, no. 1, p. 3, Dec. 2022.

[14] J. Wang and L. Wang, "Mobile edge computing task distribution and offloading algorithm based on deep reinforcement learning in Internet of Vehicles," *J. Ambient Intell. Humanized Comput.*, pp. 1–11, Aug. 2021.

[15] K. Wang, X. Wang, and X. Liu, "A high reliable computing offloading strategy using deep reinforcement learning for IoVs in edge computing," *J. Grid Comput.*, vol. 19, no. 2, pp. 1–15, Jun. 2021.

[16] K. Smida, H. Tounsi, and M. Frikha, "Intelligent and resizable control plane for software defined vehicular network: A deep reinforcement learning approach," *Telecommun. Syst.*, vol. 79, pp. 163–180, Jan. 2022.

[17] F. R. Shahidani, A. Ghasemi, A. T. Haghighat, and A. Keshavarzi, "Task scheduling in edge-fog-cloud architecture: A multi-objective load balancing approach using reinforcement learning algorithm," *Computing*, vol. 105, no. 6, pp. 1337–1359, Jun. 2023.

[18] G. Park, W. Lee, and I. Joe, "Network resource optimization with reinforcement learning for low power wide area networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2020, no. 1, pp. 1–20, Dec. 2020.
[19] M. A. Gava, H. R. O. Rocha, M. J. Faber, M. E. V. Segatto, H. Wörtche, and J. A. L. Silva, "Optimizing resources and increasing the coverage of Internet-of-Things (IoT) networks: An approach based on LoRaWAN," *Sensors*, vol. 23, no. 3, p. 1239, Jan. 2023.
[20] X. Liao, J. Shi, Z. Li, L. Zhang, and B. Xia, "A model-driven deep reinforcement learning heuristic algorithm for resource allocation in ultra-dense cellular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 983–997, Jan. 2020.
[21] S. Zhang, C. Wang, J. Zhang, Y. Duan, X. You, and P. Zhang, "Network resource allocation strategy based on deep reinforcement learning," *IEEE Open J. Comput. Soc.*, vol. 1, pp. 86–94, 2020.
[22] N. Zhao, Y.-C. Liang, D. Niyato, Y. Pei, M. Wu, and Y. Jiang, "Deep reinforcement learning for user association and resource allocation in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5141–5152, Nov. 2019.
[23] M. Jouhari, K. Ibrahimi, J. Ben Othman, and E. M. Amhoud, "Deep reinforcement learning-based energy efficiency optimization for flying LoRa gateways," 2023, *arXiv:2302.05214*.
[24] S. U. Minhaj, A. Mahmood, S. F. Abedin, S. A. Hassan, M. T. Bhatti, S. H. Ali, and M. Gidlund, "Intelligent resource allocation in LoRaWAN using machine learning techniques," *IEEE Access*, vol. 11, pp. 10092–10106, 2023.
[25] M. Ashawa, O. Douglas, J. Osamor, and R. Jackie, "Improving cloud efficiency through optimized resource allocation technique for load balancing using LSTM machine learning algorithm," *J. Cloud Comput.*, vol. 11, no. 1, pp. 1–17, Dec. 2022.
[26] M. Dehghani, Z. Montazeri, E. Trojovská, and P. Trojovský, "Coati optimization algorithm: A new bio-inspired metaheuristic algorithm for solving optimization problems," *Knowl.-Based Syst.*, vol. 259, Jan. 2023, Art. no. 110011.
[27] M. Azizi, U. Aickelin, H. A. Khorshidi, and M. B. Shishehgarkhaneh, "Energy valley optimizer: A novel metaheuristic algorithm for global and engineering optimization," *Sci. Rep.*, vol. 13, no. 1, p. 226, Jan. 2023.
[28] V. Muthiah-Nakarajan and M. M. Noel, "Galactic swarm optimization: A new global optimization metaheuristic inspired by galactic motion," *Appl. Soft Comput.*, vol. 38, pp. 771–787, Jan. 2016.
[29] A. R. Moazzeni and E. Khamehchi, "Rain optimization algorithm (ROA): A new metaheuristic method for drilling optimization solutions," *J. Petroleum Sci. Eng.*, vol. 195, Dec. 2020, Art. no. 107512.

M. RAJESWARA RAO received the B.Tech. degree in electronics and communications engineering from JNTU University, Hyderabad, Andhra Pradesh, India, in 2006, and the M.Tech. degree in VLSI systems design from JNTUA Anantapur, Andhra Pradesh, India, in 2009. He is currently pursuing the Ph.D. degree with the School of Electronics Engineering (SENSE), Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India. His research interests include deep learning, AI/ML, embedded systems, and the Internet of Things (IoT). His current research interest includes the optimization of resource allocation on LPWAN networks.

S. SUNDAR received the B.E. degree in electronics and communication engineering from the University of Madras, the master's degree in applied electronics from Anna University, and the Ph.D. degree from the Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India. Since 2010, he has been with the Department of Electronics Engineering, School of Electronics Engineering, VIT University. He has published research works in international journals. His research interests include wireless sensor networks, embedded systems, the IoT, machine learning, and artificial intelligence.