IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 October 2024, accepted 16 October 2024, date of publication 31 October 2024, date of current version 20 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3489395

RESEARCH ARTICLE

Optimized Reinforcement Learning for Resource Allocation in Vehicular Ad Hoc Networks

SPANDANA MANDE[1], NANDHAKUMAR RAMACHANDRAN[2], SHAIK SALMA ASIYA BEGUM[3], AND FERNANDO MOREIRA[4,5]

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Andhra Pradesh 522302, India
[2]SCOPE, VIT-AP University, Amaravati, Andhra Pradesh 522237, India
[3]Department of Computer Science and Engineering (AI&ML), Lakireddy Bali Reddy College of Engineering, Mylavaram, Andhra Pradesh 521230, India
[4]REMIT, IJP, Universidade Portucalense, 4200-072 Porto, Portugal
[5]IEETA, Universidade de Aveiro, 3810-193 Aveiro, Portugal

Corresponding author: Fernando Moreira (fmoreira@upt.pt)

This work was supported by the FCT–Fundação para a Ciência e a Tecnologia, I.P., under Project UIDB/05105/2020.

ABSTRACT The vehicular ad hoc networks VANET is an essential part of intelligent transportation systems (ITSs) since it may offer various multimedia services and safety services to pedestrians, passengers, and even drivers. A wireless communication protocol called dedicated short-range communication (DSRC) was created for toll collection systems. Nevertheless, DSRC standards are extremely constrained, necessitating the development of next-generation communication protocols appropriate for VANET. Here, intended to develop an Optimized Reinforcement Learning (ORL) for obtaining resource allocation in VANET. This proposed methodology is developed for achieving resource allocation with efficient data transmission. This proposed approach is utilized to adjust the control channel interval (CCI) and service channel interval (SCI) to empower network performance. Additionally, it is utilized to reduce data collisions and optimize the network's backoff distribution. The proposed method is a combination of reinforcement learning (RL) and adaptive coati optimization (ACO). The coati optimization mimics the characteristics of coati in natures in which it depends upon the coati escape from predators and hunting and attacking behaviour at various climates.The RL is utilized to obtain an efficient channel access algorithm. In the RL, the Q value is optimally selected by using ACO. Based on this algorithm, the proposed method is utilized to enhance the performance of VANET data transmission by achieving optimal resource allocation. The proposed method is implemented in MATLAB, and performances are evaluated using performance measures. Additionally, to validate the performance of the proposed methodology, it is compared with conventional techniques.

INDEX TERMS VANET, resource allocation, adaptive coati optimization and reinforcement learning.

I. INTRODUCTION

Vehicular ad hoc networks (VANETs) are inexpensive, self-organizing internet connections to share information and connect Road Side Units (RSUs) without the need for facilities [1]. Each automobile in a VANET is outfitted with portable gadgets; the innovation employs the essential connectivity modules [2]. Vehicles are equipped with an on board unit (OBU), which is made up of a primary control device, an operator panel, a specific connector for connecting to a secondary OBU [3], and storage for comprehension and composition data that has been acquired. Given that vehicles can interact with other vehicles and with roadway base stations, which are positioned at strategic junctures along the route, the situation can be characterized as an autonomous element of the transportation system.

It is composed of several devices, each of which can be an automobile or an emergency device. In VANET, there are primarily two forms of communication: vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I). It is necessary to transition between networks since vehicles are portable gadgets that happen to be in progress. Vehicle-to-vehicle (V2V) [4], refers to the inter-vehicle use of informatics and vehicle-to-everything (V2X), technologies. In this scenario,

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

automobiles form an IoT (Internet of Things) on automobiles by connecting with their peers over 5G or designated short-range connectivity (DSRC). Transportation facilities and automobiles can wirelessly share data through a system called vehicle-to-infrastructure (V2I) transmission [5]. It makes it possible for cars to transfer, endure, and transfer messages over an electronic grid structure, even erratic operators will be able to control their driving with sufficient intervals for these sensors to gather congestion a distance in advance of the vehicle.

The technique of resource allocation involves determining all of the resources that are accessible for a given task, including money and manpower, and then wisely distributing these individuals to activities that allow users to perform at their highest level [6]. The processes used to distribute items and amenities are known as distribution of resources mechanisms. The two primary categories of resource allocation under consideration are one-time and recurring [7]. Optimizing the use of restricted assets and obtaining the greatest possible return on expenditures, involves navigating conflicting requirements and priorities and choosing the most suitable course of strategy. Investing appropriate resources in the proper involvement at the appropriate moment increases productivity and reduces excess hazards. The task profitability and timeliness are maintained through the integration of connection accounts and greater insight throughout every component.

Reliability is increased and the chance of extensions is reduced when the appropriate resources are allocated to the correct activity at the proper timeframe. Increased resource availability along with involvement in economics contributes to programs staying sustainable and on track. One of VANET's primary responsibilities is to enable wireless data transfer in automobile regions so that vehicle velocity and location statistics can be sent to one another via a wireless connection, thereby preventing accidents [8]. Based on the road structure, vehicle movement, destination and controllers, the authors [24] handled dynamic variation in IoV. Due to the ad hoc based distance vector routing, an optimized path algorithm is utilized to ensure the latency and security. The IoV communication framework entered and prevented suspicious vehicles. In [24], concerned RSU short range with the all contributing vehicles authentication process agreed with RSUs and EVs. The signatures repeated verification highly decreasing and limited resource-RSUs with reducing big overhead. The revocation, edge-computing and batch verification supported with the quantum-secure protocol. The traffic-related information is fake that stopped and the malicious vehicle traced effectively. The flexibility of automobile accessibility, numerous link errors, regular shifts in the network structure, and a decentralized structure with multiple devices are the main problems with the transmission of multimedia on VANET. This work presented a novel an optimized reinforcement learning for resource allocation in VANET. Following points summarizes the major research contributions.

☐ To achieve effective VANET resource allocation using proposed RL and ACO algorithm also improving the efficiency of data transmission.

☐ To improve network performance by adjusting the parameters of SCI as well as CCI. Within the network, the data collisions minimized.

☐ To use proposed RL and ACO algorithm for optimizing the backoff distribution of network thereby developing effective channel access model.

☐ To compare the proposed ORL efficiency with previous traditional approaches thereby validating proposed ORL efficiency of resource allocation.

Rest of the work is arranged like that: Section II summarizes the literature reviews of resource allocation in VANET. The proposed framework is described in Section III followed by its experimental results are discussed in Section IV. Section V concludes the whole work.

TABLE 1. List of main acronyms and symbols.

Acronyms/Symbols	Description
VANETs	Vehicular Ad Hoc Networks
RL	Reinforcement Learning
ACO	Adaptive Coati Optimization Systems
SCI	Service Channel Interval
CCI	Control Channel Interval
ORL	Optimized Reinforcement Learning
QoS	Quality of Service
RSU	Roadside Units
OBU	On-Board Unit
DSRC	Dedicated Short-Range Communication
DSRC	Dedicated Short-Range Communication
IoT	Internet of Things
V2V	Vehicle-to-Vehicle Communication
V2I	Vehicle-to-Infrastructure Communication
V2X	Vehicle-to-Everything Communication
SINR	Signal-to-Interference-Plus-Noise-Ratio
PDR	Packet Delivery Ratio
CW	Contention Window

II. LITERATURE SURVEY

To elevate Internet of Vehicles (IoV) resource allocation complication Liang et al. [9] presented a reinforcement learning algorithm to manage flexible computing, databases, and network resources. The investigation deals with resource distribution in the IoV as a semi-Markov decision process, including an additional resource allocation device and a resource reserve technique. A hierarchical architecture and Hierarchical Reinforcement Learning (HRL) are employed to determine the best resource allocation strategies. It enhances the consumption of resources and improves individual Quality of Experience (QoE). However, it doesn't thoroughly examine how scalable massive systems can be made.

To take the computing task's latency into account, Liu et al. [10] developed a vehicle-assisted offloading scheme where vehicle servers as mobile edge servers for User equipment (UE). To tackle unpredictable congestion and inconsistent interaction circumstances, the issue is restructured into a semi-Markov process. Utilizing Q-learning and Deep Reinforcement Learning algorithms enhances the distribution of resources and choice-making computational systems. It offers an extensible, sturdy, dependable, and adaptable way to delegate work and successfully handle resources. Therefore, the primary technological issue arises from deciding which calculation position should be unloaded.

The significant transmission delay and limited computing power of the present vehicular network structures provide problems for intelligent vehicular services, Wu et al, [11] described vehicle-to-multi-edges (V2Es) as an interaction structure for connecting vehicular networks. The multi-agent reinforcement learning (MRL) approach is used to acquire an unpredictable connectivity state between vehicles and border nodes to reach selections about interface storage and task delegation. It can efficiently and rapidly implement planning strategies while lowering standard service duration by over 10%. However, this method has a flaw in that it ignores the varied significance while reusing with identical frequency.

The user interface for automobiles in Internet of Vehicles (IoV) scenarios enhances and reduces bandwidth in services; Lee et al. [12] designed a vehicular fog computing (VFC) resource allocation method that utilizes the consumption of reinforcement learning (RL) in addition to a heuristic algorithm. It improves the vehicle's computing situations, safety and accident alerts, driveway status, problem-solving, and customer satisfaction rates. It improves connected devices' overall standard of experience in modern cities. Thus, it is not stated very great-scale transportation systems with an enormous amount of simultaneous vehicles.

The analysis uses deep reinforcement learning for flexible spectrum availability and Wang et al. [13] suggest a multi-hop broadcast protocol (G-hop) to handle the problem of scarce frequency assets in vehicular ad hoc networks (VANETs). Using the depth-first-search technique classifies vehicles according to attributes like transmission distance and acceleration of travel. To comprehend the time-dependent procedure, it presents the Global Optimization method based on Experience Accumulation (GOEA), which integrates an interconnected framework utilizing deep Q-network and recurrent neural networks. Both methods are reliable and efficient in VANETs through frequency assents and multi-hop communication. Thus, it does not examine any potential issues implemented in in-vehicle networks.

Using reinforcement learning in Vehicular Ad Hoc Networks (VANETs) to improve transmission effectiveness and efficacy Luo et al. [14] suggest an intersection-based V2X Routing algorithm to implement Q-learning at crossings and using Roadside Units (RSUs) for optimal navigation. SUMO (Simulation of Urban Mobility) traffic software which is frequently used to investigate transportation networks and VANET situations. The effectiveness and flexibility of changing roadway circumstances are assessed through the creation of VANET settings with several crossings and moving automobiles within the framework. Hence, the effect of changing network settings or outside factors is not considered.

To maximize resource allocation selections, Liang et al. [15] use a Long Short-Term Memory (LSTM) neural network in a deep learning-based model for transportation networks. To evaluate the efficacy and precision of the algorithm for assigning resources to automobiles depending on current network circumstances and specifications, the investigation also carries out performance tests. The technique blends moving vehicle statistics with the system's current states to enhance the effectiveness of resource allocation and enable flexible choice-making in volatile vehicular contexts. Therefore, it is inadequate to evaluate for various network conditions.

The efficiency of such programs can be improved by vehicular networking (VN) technology, which gives smart vehicle (SVs) connectivity to the computational and detecting power of nearby SVs. Due to their portability and the characteristics of transmitted data routes, Chattopadhyay and Tham [16] develop Contextual Bandits (CB), Markov decision process (MDP), and deep Q-network (DQN) based methods for VN detection and analysis. Therefore, improving SV systems' reliability can be achieved through the development of effective computational and detecting strategies for SVs and VNs. Thus, it is not always dependable.

Because vehicle settings are dynamic, it makes sense to develop a hybrid technique for efficient resource allocation. Kumar et al. [17] utilize deep learning methods to forecast the movement patterns of automobiles. Then, to optimize the bandwidth performance of all participating vehicles, by utilizing two deep reinforcement learning methods: advantage actor-critic (A2C) and deep Q-network (DQN). Considering the temporal variability of user flexibility, it integrates the long short-term memory (LSTM) into DQN and A2C approaches. It demonstrates the efficacy and reliability of the suggested approaches.

III. PROPOSED MODEL

The proposed workflow model is outlined in Figure 1. The service announcement of wireless access in vehicular environment incorporating update time information, guard interval (guard interval length information), CCI and SCI are the channel adjustment parameters that included in RSUs. The control channel monitored continuously via OBUs in vehicles before connecting to a particular RSU. The RSU with OBU performing synchronization process after obtaining the advertisement message. Due to channel adjustment parameters, the channel switching among service and control channels. To perform channel switching, the service and control channel accomplishing data exchanging via OBUs and RSUs in vehicle. Due to enlarger data traffic, the present

CCI and SCI congestion level monitored via RSUs during data exchange. More crucial is CCI because of management message transmissions such as high-priority messages, timing information and service advertisements. From RSU to OBU, transmit the parameters of adapted channel coordination, which adjust the channel switching. The CCI and SCI adjustment continued data exchange. The data exchange efficiency with the number of data traffic in CCI and SCI congestion levels are monitored.

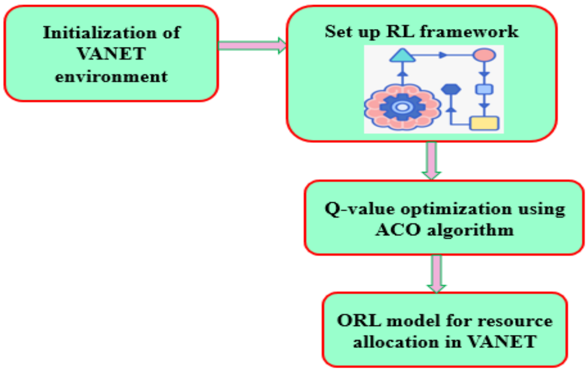

FIGURE 1. **Proposed workflow model for resource allocation in VANET.**

A. REINFORCEMENT LEARNING MODEL

To provide better resource allocation, this work proposed a model of Reinforcement Learning (RL) model thereby proving effective data transmission in VANET [18]. An optimal strategies learned with enabling the evaluations of critic (function) and actor (policy) combination. Figure 2 shows the framework of Actor-Critic based on RL.

1) ACTOR-CRITIC MODEL

An effective learning and stable procedure allowed via Actor-Critic outline. While making decisions and detecting optimal policy, the major responsible is Actor network in which the feedback on those decisions is provided with the help of Critic network [19]. The resource allocation performed by ensuring adaptability. Below expression mathematically representing the framework of Actor-Critic.

$$\pi\left(b \backslash d;\theta^{\pi}\right)\approx\pi^{*}\left(b \backslash d\right)$$
$$Q\left(d,b;\theta^{Q}\right)\approx Q^{*}\left(d,b\right) \tag{1}$$

Based on the above equation, d as an Actor mapping state to action as b. Critic computes the function of action-value is $Q\left(d,b;\theta^{Q}\right)$. Both Critic and Actor performance improved over training is updated with θ^{Q} and θ^{π} parameters.

2) ACTOR NETWORK

Due to the current state of VNAET, the decision on resource allocation made with Actor network. The available actions based on probability distribution created with θ^{π} adapts $\pi\left(b \backslash d,\theta^{\pi}\right)$ as the policy function. The resource allocation efficiency raised with expected cumulative reward increased and the actions are selected that involved in decision making process.

$$b_t\approx\pi\left(b \backslash d;\theta^{\pi}\right) \tag{2}$$

3) CRITIC NETWORK

Here, is the action value function that resembled the Actor's decision efficacy evaluated with Critic network [20]. The process of Actor's learning guided to provide feedback also the Actor in the given state takes action values are evaluated

$$Q\left(d,b;\theta^{Q}\right)\approx R_t+\delta Q\left(d_{t+1},b_{t+1};\theta^{Q}\right) \tag{3}$$

In the state d_t, obtain an immediate reward is R_t. Based on future rewards, the discount factor is δ.

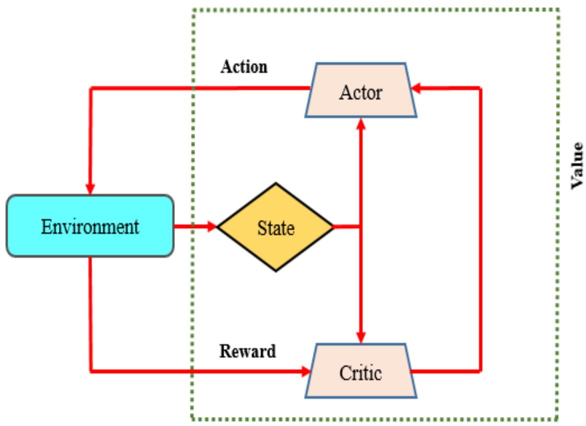

FIGURE 2. **Framework of RL based Actor-Critic model.**

B. ADAPTIVE COATI OPTIMIZATION (ACO)

The Q-values in RL is optimized via ACO algorithm. The latency reduced with effective adapt transmission with the service channel and control utilizations optimized via ACO [21]. The coati optimization mimics the characteristics of coati in natures in which its depends upon the coati escape from predators and hunting and attacking behaviour at various climates. This optimization is driven based on the coatis and the location of coati is utilized for the determination of decision variables. The initial locations are driven randomly in the search space and is framed as,

$$G_j: G_{j,i}=lc_i+m.\left(uc_i-lc_i\right),\ \ j=1,2,\ldots,K,$$
$$i=1,2,\ldots,l \tag{4}$$

The total number of coatis is K with the decision variables l. the j^{th} coati in the search dimension is $G_{j,i}$. The random number is m and is falls under the range of 0, 1. Based on the time, an optimal channel learned with effective channel access model developed via utilizing better Q-values in RL. The bottommost and uppermost constraints of the search area is lc and uc. Improve the resource allocation based on intelligent decision making with network environment changing adapted with RL model. The matrix form of the COA in the population matrix is framed as,

$$G=\begin{bmatrix} G_1 \\ \vdots \\ G_j \\ \vdots \\ G_K \end{bmatrix}=\begin{bmatrix} g_{1,1} & \cdots & g_{1,i} & \cdots & g_{1,k} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ g_{j,1} & \cdots & g_{j,i} & \cdots & g_{j,l} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ g_{K,1} & \cdots & g_{K,i} & \cdots & g_{K,l} \end{bmatrix} \tag{5}$$

The positioning of the best solutions with respect to the decision variables are responsible for the objective function issue as Q-value optimization and can be formulated as,

$$O=\begin{bmatrix} O_1 \\ \vdots \\ O_j \\ \vdots \\ O_K \end{bmatrix}_{K\times 1}=\begin{bmatrix} O(G_1) \\ \vdots \\ O(G_j) \\ \vdots \\ O(G_K) \end{bmatrix}_{K\times 1} \tag{6}$$

The objective function of the Coati is O and j^{th} coati's objective function is O_j [22]. The location up-gradation of the coatis is based on two strategies, which is explained detailed in the following sub-section.

1) ATTACKING THE IGUANAS

$$G_j^{S1}: G_{j,i}^{S1}=g_{j,i}+m.\left(Igu_i-F.g_{j,i}\right),\quad j=1,2,\ldots,\frac{K}{2}, \\ i=1,2,\ldots,l \tag{7}$$

After the attack the iguana might have fall on the ground and hence the location of the coati changes with that. This can be expressed as,

$$Igu^Z: Igu_i^Z=lc_i+m.\left(uc_i-lc_i\right),\quad i=1,2,\ldots,l \tag{8}$$

$$G_j^{S1}: G_{j,i}^{S1}=\begin{cases} g_{j,i}+m.(igu_i^Z-F.g_{j,i}), & O_{Igu^Z}<O_j \\ g_{j,i}+m.\left(g_{j,i}-Igu_i^Z\right), & else \end{cases} \\ for\ j=\left[\frac{K}{2}\right]+1,\ \left[\frac{K}{2}\right]+2,\ldots,K \\ and\ \ i=1,2,\ldots,l \tag{9}$$

2) ESCAPING FROM THE PREDATOR

The escaping from the predator is simulated for the random location of coati and is the second phase and can be outlined as,

$$lc_i^{local}=\frac{lc_i}{t},\ uc_i^{local}=\frac{uc_i}{t},\ where\ t=1,2,\ldots,T \tag{10}$$

$$G_j^{S1}: G_{j,i}^{S1}=g_{j,i}+(1-2m) \\ .\left(lc_i^{local}+m.\left(uc_i^{local}-lc_i^{local}\right)\right), \\ j=1,2,\ldots,K,\ i=1,2,\ldots,l \tag{11}$$

The updated location is based on the objective function and can be stimulated, which is enhanced with the help of distance vector formula thereby providing better results of ACO.

$$G_j=\begin{cases} G_j^{S2}, & O_j^{S2}<O_j \\ G_j, & else \end{cases} \tag{12}$$

G_j^{S2} is the updated location for the j^{th} coati and the i^{th} dimension of G_j^{S2} and the respective objective function is O_j^{S2}. The random number with the duration of 0 and 1 and the iteration counter is t. the local bottommost and uppermost constraints of the decision variable are lc_i^{local} and uc_i^{local}.

$$D(y,u)=\sqrt{\sum_{t=1}^{d}\left(\frac{Y_j-U_j}{y_{\max}-y_{\min}}\right)^2} \tag{13}$$

where y and u are the two vector distance is $D(y,u)$. The minimum and maximum vector values of Y is $y_{\max}$ and $y_{\min}$. Based on RL framework, the Q-values are optimized by introducing ACO algorithm and more effective actions are selected. This model provided better performances with faster convergence as well as an optimized Q-values provided an enhanced learning process.

C. THE ORL MODEL FOR RESOURCE ALLOCATION IN VANET

The network efficiency enhanced with the resource allocation in VANETs using Reinforcement Learning and ACO algorithm. For real-time resource management, the adaptability of ACO and strategic decision-making accomplished with RL. An average network load, traffic patterns, typical road conditions based channel prioritization included in RL. The long-term strategies optimized using ACO and to anticipate resource needs.

Initialize the VANET parameters and nodes. Q-value is optimized with the usage of ACO algorithm it boosts up the RL performance more effectively. The ORL training process directing is the objective function along with better resource allocation. Optimize the metric severed as the objective function, which is Q-value [25], [26], [27], [28], [29]. The Critic validation and Actor's decision making procedure essence captured with components. Based on time, the expected cumulative reward maximization as the actor aim. Following expression represents the Actor objective function ($I\left(\theta^{\pi}\right)$). Network conditions like traffic density, channel condition and vehicle position data collected. The Q-value is selected and optimized using ACO. The optimized SCI and CCI adjusted via Q-value.

$$I\left(\theta^{\pi}\right)=F_{\pi}\left[\sum_{t=0}^{\infty}\delta^t R_t\right] \tag{14}$$

where, $\pi\left(b\backslash d;\theta^{\pi}\right)$ is the policy takes expectations with action, immediate reward action and reward importance discount factor is represented as b_t, R_t and δ. The aim of Critic in QRL is to adjust both SCI and CCI parameters thereby enhancing the performance of network. Following formula shows the Critic objective function ($I\left(\theta^{\pi}\right)$).

$$I\left(\theta^{\pi}\right) \\ =F\left[\left(Q\left(d_t,b_t;\theta^Q\right)-\left(R_t+\delta Q\left(d_{t+1},b_{t+1};\theta^Q\right)\right)\right)^2\right] \tag{15}$$

The decision-making enhanced with Critic's feedback and a cooperative learning process ensured via joint function based on Critic's objective. More accurate evaluation is provided with value function of Critic refines and the higher expected cumulative rewards led via actions pursued from Actor. The selected channel access strategy related to data transmission performed.

$$I\left(\theta^{\pi},\theta^{Q}\right)=I\left(\theta^{\pi}\right)-\alpha I\left(\theta^{Q}\right) \tag{16}$$

The trade-off among minimizing difference error and maximizing expected reward are balanced with the help of hyper parameter α. Record and observe backoff distribution with data collisions. Where, θ^{Q} and θ^{π} are the parameters during the training for joint objective function optimization. The resource allocation in enhanced and optimal Q-values sets are involved in training process. The actions and reward observes the parameters are iteratively updated in the process of training. Reward calculated according to energy efficiency, rate of collision and data transmission with better resource allocation provided. Proposed ORL model for resource allocation is outlined in Figure 3.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The proposed technique utilizes the of Simulation of MATLAB software there by showing better performances of proposed framework. The traffic parameters used in this study are enlisted in Table 2.

TABLE 2. Traffic simulation parameters.

Factors	Range
No.of Vehicles	30-110
Safety Packet Size	256 bytes
Packet generation frequency	10 Hz
Channel bandwidth	10MHz
Propagation Model	Nakagami Model
Transmission Power	30dBM
Channel Frequency	5.87 GHz
Non Safety Packet Size	400 bytes

A. PERFORMANCE EVALUATION

The performance evaluation of proposed resource allocation in VANET follows four parameters namely, throughput, packet delivery ratio, delay (ms), and Collision rate (%). The performance of the proposed work is compared with the existing techniques known as HRL [9], QDRL [10], MRL [11], and DQRNN [13]. Firstly, the throughput of the proposed and comparative study is deliberately explained in Figure 4 of graphical representation. Throughput in the VANET is the estimation of how much data transmitted from

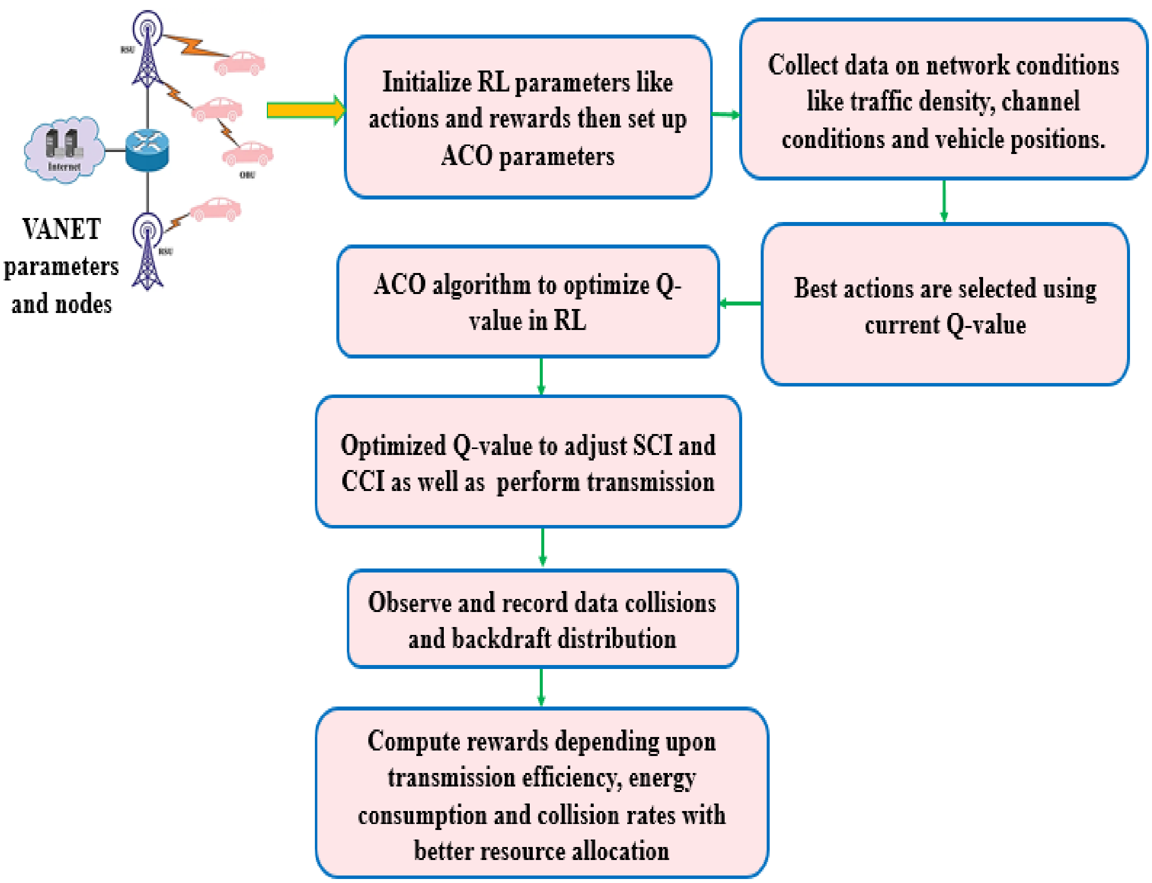

FIGURE 3. Proposed ORL model for resource allocation.

the source to destination at the given time without any delay. Since VANET is dynamic in nature the resource allocation is difficult followed by lower throughput. Sometimes, the vehicles inside the network changes its direction or slow down or even faster and the throughput might be reduced with struggled resource allocation. The proposed approach robustly maintains the issues and the throughput is higher when compared to the other techniques such as HRL [9], QDRL [10], MRL [11], and DQRNN [13]. The throughput is measured with the number of vehicles and it got reduced with the increasing number of vehicles as shown in figure 4. The throughput of proposed technique when the number of vehicles are 100 is 8.9Mbps. the other techniques such as HRL [9], QDRL [10], MRL [11], and DQRNN [13] achieved throughput of 5.8Mbps, 6.4 Mbps, 6.1 Mbps, and 7.0Mbps correspondingly.

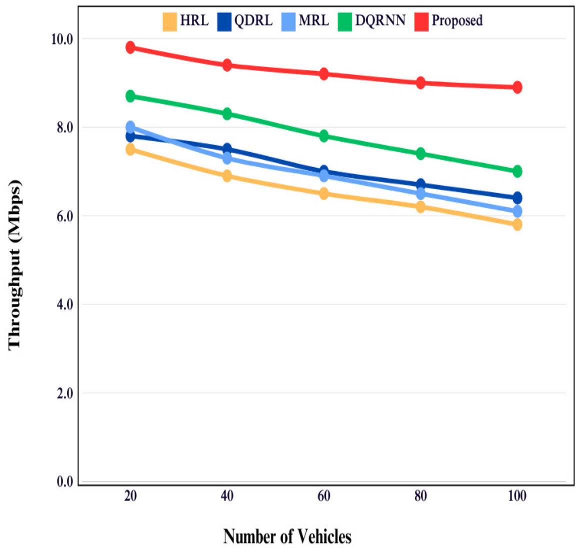

FIGURE 4. Throughput analysis with respect to the number of vehicles.

PDR or Packet Delivery ratio is measured as the data transmitted from one source vehicle to the destination vehicle through other intermediate vehicles without any loss. Higher the PDR represents higher the reliability and efficiency. It also responsible for the resource allocation and routing protocols of the VANET networks. The graphical analysis of the PDR with respect to the number of vehicles is shown in Figure 5. From the figure, it is observed, the PDR of the proposed work is higher with 9 value and other techniques such as HRL [9], QDRL [10], MRL [11], and DQRNN [13] achieved 7.0, 7.3, 7.8, and 8.2 respectively. this ensures the reliability of the proposed technique when compared to the other techniques.

Delay is measured as the time taken by the vehicles to transmit the data from the source vehicle to the destination vehicle. This purely depends on the resources that are allocated to the vehicles inside the network. Lower delay determines higher the reliability. The graphical representation of the proposed work's delay and other techniques such as HRL [9], QDRL [10], MRL [11], and DQRNN [13] are visualized in Figure 6. The delay of the proposed technique

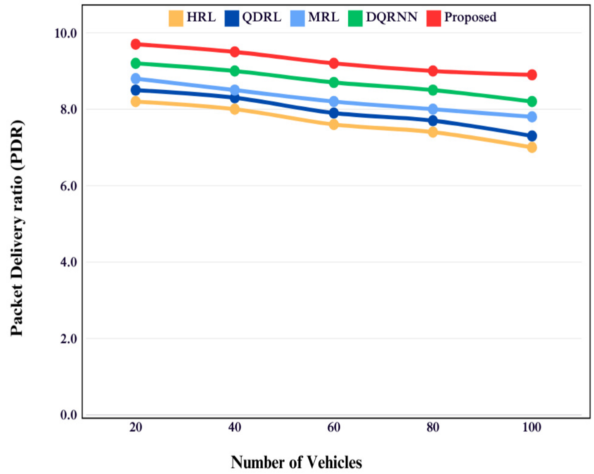

FIGURE 5. PDR analysis with respect to the number of vehicles.

is lower than the other techniques, however, it increases with the increasing number of vehicles. The delay of the proposed work is 4ms when the number of vehicles are equal to 100 and other techniques showed the delay of 7.8, 6.9, 8.1, and 7.3 ms.

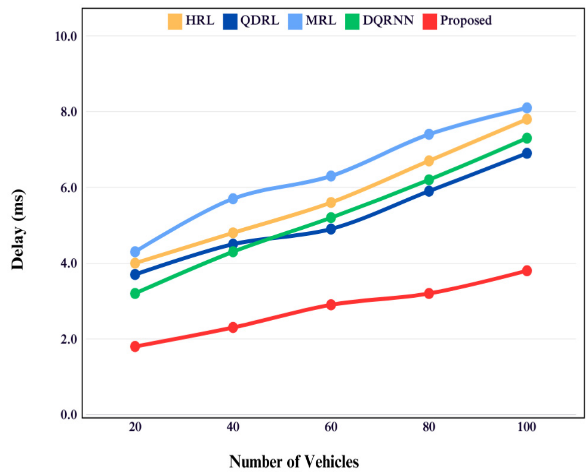

FIGURE 6. Delay with respect to the number of vehicles.

Figure 7 illustrates the graphical illustrations of the collision rate with respect to the number of nodes. The collision rate of proposed work is lower than the other techniques such as HRL [9], QDRL [10], MRL [11], and DQRNN [13]. The tolerance capacity of the proposed work in terms of channel intervals and CW is higher and hence mitigates the collision of vehicles. The collision rate increases with the increasing number of vehicles and is 6% when the number of vehicles are 100 as shown in figure. Other techniques such as HRL [9], QDRL [10], MRL [11], and DQRNN [13] showed the collision rate of 40.2, 34.5, 32.7, and 25.6 correspondingly.

The comparative tabulation for overall execution time is outlined in Table 3. The state-of-art like HRL [9], QDRL [10], MRL [11], and DQRNN [13] with proposed approach is used for overall execution time analysis. The proposed model outperformed 3.4s whereas HRL [9], QDRL [10], MRL [11],

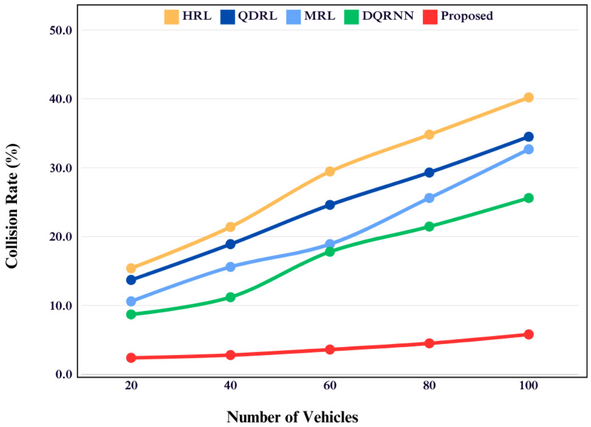

FIGURE 7. Collision rate with respect to the number of vehicles.

and DQRNN [13] provided 13.23s, 27.41s, 14.53s and 9.82s of execution time. Compared to all other existing methodologies, the proposed model's effectiveness in VANET becomes lower execution time.

TABLE 3. State-of-art comparison for overall execution time.

Methods	Execution time (Sec)
Proposed	3.4
HRL [9]	13.23
QDRL [10]	27.41
MRL [11]	14.53
DQRNN [13]	9.82

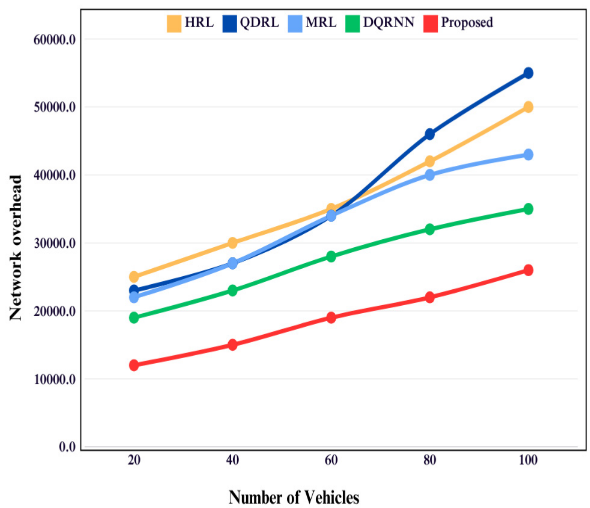

FIGURE 8. Network overhead with respect to the number of vehicles.

Figure 8, illustrates the visual diagram of the proposed and other techniques such as HRL [9], QDRL [10], MRL [11], and DQRNN [13] with respect to network overhead and number of vehicles. The network overhead of proposed work when the number of vehicles equal to 100 is 26,000 and other techniques HRL [9], QDRL [10], MRL [11], and DQRNN [13] showed network overhead of 50,000, 55,000, 43,000 and 35,000 respectively as shown in Figure 8. The proposed work surpasses all the other techniques.

V. CONCLUSION

Multiple automobiles link autonomously to the network through wireless technologies to form VANET. DSRC routinely transfers data at speeds of more than several hundred kbps over the 5 GHz range. This work presented a novel optimized reinforcement learning for resource allocation in VANET. Due to an optimized Q-values, real-time adjustments via network performance enhanced. Minimizing collisions and enhancing the efficiency of data transmission. Backoff time optimized with network resource ensured with more effective. Based on channel access strategies, more efficient decision making by utilizing optimized Q-value with the help of ACO algorithm. MATLAB software handles the experimental implementation and reveals the efficiency of proposed with comparative approaches. Compared with existing methods, the experimental outputs demonstrated enhanced results in the efficiency of data transmission and resource allocation. According to real-word vehicular networks, VANET performances enhanced and efficacy of ORL model validates implementation and its shows the potential to practical applications. When compared with existing HRL, QDRL, MRL and DQRNN, the proposed method outperformed 4ms delay, 9% of PDR, 8.9Mbps throughput and 6% rate of collision. The proposed method met the challenges of computational complexity, scalability, communication overhead and hardware limitations.

REFERENCES

[1] M. A. Al-Shareeda and S. Manickam, "A systematic literature review on security of vehicular ad-hoc network (VANET) based on VEINS framework," *IEEE Access*, vol. 11, pp. 46218–46228, 2023.

[2] D. Abada, A. Massaq, A. Boulouz, and M. B. Salah, "An adaptive vehicular relay and gateway selection scheme for connecting VANETs to internet via 4G LTE cellular network," in *Emerging Technologies for Connected Internet of Vehicles and Intelligent Transportation System Networks* (Studies in Systems, Decision and Control), vol. 242, M. Elhoseny and A. Hassanien, Eds., Cham, Switzerland: Springer, 2020, doi: 10.1007/978-3-030-22773-9_10.

[3] J. Santa, L. Bernal-Escobedo, and R. Sanchez-Iborra, "On-board unit to connect personal mobility vehicles to the IoT," *Proc. Comput. Sci.*, vol. 175, pp. 173–180, Jan. 2020.

[4] M. Shurrab, S. Singh, H. Otrok, R. Mizouni, V. Khadkikar, and H. Zeineldin, "An efficient vehicle-to-vehicle (V2V) energy sharing framework," *IEEE Internet Things J.*, vol. 9, no. 7, pp. 5315–5328, Apr. 2022.

[5] P. Sun, D. Nam, R. Jayakrishnan, and W. Jin, "An eco-driving algorithm based on vehicle to infrastructure (V2I) communications for signalized intersections," *Transp. Res. C, Emerg. Technol.*, vol. 144, Nov. 2022, Art. no. 103876.

[6] M. Noor-A-Rahim, Z. Liu, H. Lee, G. G. M. N. Ali, D. Pesch, and P. Xiao, "A survey on resource allocation in vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 701–721, Feb. 2022.

[7] Y. Xu, G. Gui, H. Gacanin, and F. Adachi, "A survey on resource allocation for 5G heterogeneous networks: Current research, future trends, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 668–695, 2nd Quart., 2021.

[8] M. J. N. Mahi, S. Chaki, S. Ahmed, M. Biswas, M. S. Kaiser, M. S. Islam, M. Sookhak, A. Barros, and M. Whaiduzzaman, "A review on VANET research: Perspective of recent emerging technologies," *IEEE Access*, vol. 10, pp. 65760–65783, 2022.

[9] H. Liang, X. Zhang, X. Hong, Z. Zhang, M. Li, G. Hu, and F. Hou, ''Reinforcement learning enabled dynamic resource allocation in the Internet of Vehicles,'' *IEEE Trans. Ind. Informat.*, vol. 17, no. 7, pp. 4957–4967, Jul. 2021.
[10] Y. Liu, H. Yu, S. Xie, and Y. Zhang, ''Deep reinforcement learning for offloading and resource allocation in vehicle edge computing and networks,'' *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11158–11168, Nov. 2019.
[11] J. Wu, J. Wang, Q. Chen, Z. Yuan, P. Zhou, X. Wang, and C. Fu, ''Resource allocation for delay-sensitive vehicle-to-multi-edges (V2Es) communications in vehicular networks: A multi-agent deep reinforcement learning approach,'' *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1873–1886, Apr. 2021.
[12] S.-S. Lee and S. Lee, ''Resource allocation for vehicular fog computing using reinforcement learning combined with heuristic information,'' *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10450–10464, Oct. 2020.
[13] Y. Wang, X. Li, P. Wan, and R. Shao, ''Intelligent dynamic spectrum access using deep reinforcement learning for VANETs,'' *IEEE Sensors J.*, vol. 21, no. 14, pp. 15554–15563, Jul. 2021.
[14] L. Luo, L. Sheng, H. Yu, and G. Sun, ''Intersection-based V2X routing via reinforcement learning in vehicular ad hoc networks,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5446–5459, Jun. 2022.
[15] L. Liang, H. Ye, G. Yu, and G. Y. Li, ''Deep-learning-based wireless resource allocation with application to vehicular networks,'' *Proc. IEEE*, vol. 108, no. 2, pp. 341–356, Feb. 2020.
[16] R. Chattopadhyay and C.-K. Tham, ''Joint sensing and processing resource allocation in vehicular ad-hoc networks,'' *IEEE Trans. Intell. Vehicles*, vol. 8, no. 1, pp. 616–627, Jan. 2023.
[17] A. S. Kumar, L. Zhao, and X. Fernando, ''Multi-agent deep reinforcement learning-empowered channel allocation in vehicular networks,'' *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1726–1736, Feb. 2022.
[18] J.-W. Kim, J.-W. Kim, and J. Lee, ''Intelligent resource allocation scheme using reinforcement learning for efficient data transmission in VANET,'' *Sensors*, vol. 24, no. 9, p. 2753, Apr. 2024.
[19] P. Upadhyay, V. Marriboina, S. J. Goyal, S. Kumar, E.-S.-M. El-Kenawy, A. Ibrahim, A. A. Alhussan, and D. S. Khafaga, ''An improved deep reinforcement learning routing technique for collision-free VANET,'' *Sci. Rep.*, vol. 13, no. 1, p. 21796, Dec. 2023.
[20] X. Liu, B. S. Amour, and A. Jaekel, ''A reinforcement learning-based congestion control approach for V2V communication in VANET,'' *Appl. Sci.*, vol. 13, no. 6, p. 3640, Mar. 2023.
[21] F. A. Hashim, E. H. Houssein, R. R. Mostafa, A. G. Hussien, and F. Helmy, ''An efficient adaptive-mutated coati optimization algorithm for feature selection and global optimization,'' *Alexandria Eng. J.*, vol. 85, pp. 29–48, Dec. 2023.
[22] A. Qtaish, M. Braik, D. Albashish, M. T. Alshammari, A. Alreshidi, and E. J. Alreshidi, ''Enhanced coati optimization algorithm using elite opposition-based learning and adaptive search mechanism for feature selection,'' *Int. J. Mach. Learn. Cybern.*, pp. 1–34, May 2024, doi: 10.1007/s13042-024-02222-3.
[23] T. Sharma, M. R. Kumar, S. Kaushal, D. Chaudhary, and K. Saleem, ''Privacy aware post quantum secure ant colony optimization ad hoc on-demand distance vector routing in intent based Internet of Vehicles for 5G smart cities,'' *IEEE Access*, vol. 11, pp. 110391–110399, 2023.
[24] D. Dharminder, S. Kumari, and U. Kumar, ''Post quantum secure conditional privacy preserving authentication for edge based vehicular communication,'' *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 11, p. e4346, Nov. 2021.
[25] G. Sun, L. Sheng, L. Luo, and H. Yu, ''Game theoretic approach for multipriority data transmission in 5G vehicular networks,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24672–24685, Dec. 2022.
[26] G. Sun, Y. Zhang, H. Yu, X. Du, and M. Guizani, ''Intersection fog-based distributed routing for V2V communication in urban vehicular ad hoc networks,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2409–2426, Jun. 2020.
[27] G. Sun, L. Song, H. Yu, V. Chang, X. Du, and M. Guizani, ''V2V routing in a VANET based on the autoregressive integrated moving average model,'' *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 908–922, Jan. 2019.
[28] G. Sun, Y. Zhang, D. Liao, H. Yu, X. Du, and M. Guizani, ''Bus-trajectory-based street-centric routing for message delivery in urban vehicular ad hoc networks,'' *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7550–7563, Aug. 2018.
[29] G. Sun, Z. Wang, H. Su, H. Yu, B. Lei, and M. Guizani, ''Profit maximization of independent task offloading in MEC-enabled 5G Internet of Vehicles,'' *IEEE Trans. Intell. Transp. Syst.*, early access, Jun. 25, 2024, doi: 10.1109/TITS.2024.3416300.

SPANDANA MANDE received the M.C.A. degree in computer science from Acharya Nagarjuna University, in 2008, Andhra Pradesh, India, and the M.Tech. degree in computer science and technology from Andhra University, Andhra Pradesh, in 2012. She is currently Assistant Professor, Department of CSE (H2), Koneru Lakshmaiah Education Foundation, Vaddeswaram, India, pin-522302. Her research interests include wireless ad hoc and sensor networks, 6G, and machine learning.

NANDHAKUMAR RAMACHANDRAN received the B.Tech. degree in information technology and the M.E. degree in computer science and engineering from Anna University, Tamil Nadu, India. He is currently an Associate Professor with VIT-AP University. He published various papers in reputed journals and conferences. His research interests include wireless sensor networks, mobile ad hoc networks, the IoT, network security, and machine learning. He was awarded his Ph.D. from Anna University.

SHAIK SALMA ASIYA BEGUM received the B.Tech. and M.Tech. degrees in computer science and engineering from Jawaharlal Nehru Technological University Kakinada, India, and the Ph.D. degree in computer science and engineering from the School of Computer Science and Engineering, VIT-AP University, Amaravati, India. She is currently an Associate Professor with the Department of Computer Science and Engineering (AI&ML), LBRCE, India. Her research interests include machine learning, deep learning, and image processing.

FERNANDO MOREIRA received the bachelor's degree in computer science, the master's and Ph.D. degrees in electronic engineering from the Faculty of Engineering, University of Porto, in 1992, 1997, and 2003, respectively, and the Habilitation degree, in 2018. He established the Science and Technology Department and served as the Head, from May 2018 to February 2022. He has been a dedicated member of the Science and Technology Department, Universidade Portucalense, since 1992, where he currently holds the position of a Full Professor. Additionally, he is a Visiting Professor with several other esteemed universities. His teaching portfolio encompasses various subjects spanning undergraduate and postgraduate studies. He also supervises Ph.D. and master's students. He has significantly contributed to his field with more than 275 articles and peer-reviewed scientific publications in national and international journals and conferences. His primary research interests include mobile computing, ICT in higher education, mobile learning, social business, and digital transformation. He serves as a member of the editorial advisory board for various journals and books and has organized numerous special issues for JCR journals. His commitment to advancing research is reflected in his consistent participation as the conference chair, the workshop chair, the chief guest, the keynote speaker, and in the program and scientific committees for national and international conferences. He held the M.Sc. in computation coordinator position for a decade, demonstrating his dedication to academic leadership. He is a REMIT Research Center Steering Committee Member, with extensive editorial experience and co-editing several books. His professional affiliations include NSTICC, ACM, and IEEE. He received the prestigious Atlas Elsevier Award for his contributions, in April 2019.