Journal of Advances in Communication and Networking Research
Vol: 1 Issue: 2 July 2025

Efficient Resource Allocation Scheme for D2D Communications in Cellular Networks using Attention-based Deep Reinforcement Learning

[1]Satyavarapu D Sampath Sai Anudeep and [2]Shaik Sohail,
[1]Department of CSE,
GITAM Deemed to be University, Bengaluru, India
anudeepsdss@gmail.com

[2]Department of Artificial Intelligence and Machine Learning,
Faculty of Science and Technology (IcfaiTech),
ICFAI Foundation for Higher Education, Hyderabad, India
sohail0222@gmail.com

Abstract: Device-to-device (D2D) communications can create direct communication links between users to satisfy the requirement of the Base Station (BS). As compared to the BS, the transmission power of D2D is very low. By using power superposition, the spectrum efficiency is greatly increased. Therefore, this work proposes a Deep Reinforcement Learning (DRL) based approach for resource allocation on the D2D Communications Cellular Networks. The major aim of the Attention-DRL-based Resource Allocation (RA) model is to determine the spectrum channel and transmission power of the D2D link to improve the throughput of the cellular and D2D links in the communication network. The resources are allocated to the D2D links using the A-DRL according to the dynamically varying network setup. During the interaction between the devices, the efficient use of power control and resources is enabled by the proposed A-DRL, and this model is more helpful for fulfilling the client's file request with increased throughput. Finally, the proposed model is evaluated against the existing models to prove its efficacy in resource allocation.

Keywords: Efficient Resource Allocation Mechanism; Device-to-Device Communications; Attention-based Deep Reinforcement Learning; Performance Comparison

Received: June 13, 2025 Revised: June 30, 2025 Accepted: July 1, 2025

Nomenclature

Abbreviation	Definition
A-DRL	Attention-based Deep Reinforcement Learning
AI	Artificial Intelligence
BS	Base Station
CUE	Cellular User Equipment
DL	Deep Learning
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
DRL-KM	Deep Reinforcement Learning with Kuhn-Munkres algorithm
DUE	Device-to-Device User Equipment
D2D	Device-to-Device
DUER	D2D User Equipment Receiver
DUET	D2D User Equipment Transmitter
eNB	Evolved Node B
HFDCSA	Hybrid Flow Direction with Chameleon Swarm Algorithm
HPSCAV	Hybrid Particle Swarm Cauchy Approach to African Vulture Optimization
IoT	Internet of Things
MEC	Multi-access Edge Computing
ML	Machine Learning
PSO	Particle Swarm Optimization
QoS	Quality of Service
RA	Resource Allocation
RL	Reinforcement Learning
SNR	Signal Noise Ratio
UAV	Unmanned Aerial Vehicle
UE	User Equipment

1. Introduction

The growing demand for wireless and mobile users is fulfilled by the spectrum resources [6]. Due to the restricted amount of available frequency, the shortage of spectrum resources has occurred [7]. The High QoS of the user are maintained with spectrum resources via the cellular network [8]. A prospective remedy for the spectrum shortage issue was the introduction of D2D communication technology. D2D technology allows nearby mobile users to communicate with one another directly without placing a significant burden on cellular networks [9] [10]. The cellular users occupy the spectrum channel in the D2D network thus, it is compulsory to equally

distribute the resources in the D2D network for maintaining the QoS [11]. In the presence of cellular devices, the RA is a fundamental issue in the D2D network.

Numerous research projects have adopted the RA for D2D communications as the main topic. The RA can benefit from the use of effective one-to-one mapping approaches when various D2D linkages occupy different cellular channels [12]. When several D2D are dispersed densely within a cell, the problem of RA is not solved analytically [13]. Numerous DL-based RA systems have been developed in previous works. Some of the researchers have maximized the sum rate of the D2D links using the hybrid power allocation; also it solves the minimum rate restriction, while a hybrid power allocation scheme was proposed to maximize the sum rate of D2D users while mitigating the violation of the QoS constraints [14]. However, the computational complexity of the centralized single-agent RA is high while improving the QoS [15]. However, decentralized multi-agent RA schemes have a lower QoS due to their reduced computational requirements. So, this research work developed an efficient RA scheme using DRL on the D2D network. Recently, blockchain-based MEC has also been proposed for enhancing trust and processing efficiency in wireless environments [16]. Likewise, AI-driven IoT healthcare systems highlight the importance of intelligent resource control in latency-sensitive communication [17]. Moreover, federated learning in IoT networks with adaptive strategies shows relevance for DRL-based allocation in distributed settings [18]. Similarly, DL-assisted intelligent vehicle communication systems deal with interference and reliability challenges in dense communication scenarios [19]. Furthermore, UAV path planning through swarm-based optimization highlights how dynamic trajectory and power control are interconnected in wireless networks [20]. The comparative analysis of the existing models is given in Table I.

Table 1. Comparative analysis of the prior models

Author [citation]	Methodology	Features	Challenges
Li and Guifenchen [1]	Dcdd-MCTS	• The transmit power of the user is solved by this model.	• The large-scale collaborative network is not handled by this model.
Yu and Tang [2]	DRL-KM	• It only uses the local information for optimizing the spectrum power allocation within the network.	• The effects of user density on the cooperative network are not handled by this model.
Pasha and Mohamad [3]	HPSCAV	• The QoS constraint of the user is fulfilled by this model.	• The energy consumption of the model is high.
Chennaboin and. Nandakumar [4]	HFDCSA	• The high consumption of energy is greatly solved using this model and the link rate of the communication network is reduced by this model.	• The convergence rate of the algorithm is very low.
Pan and Jieyang [5]	DQN	• The energy efficiency and network stability can be greatly enhanced using this model.	• The adaptability of the model is very low.

The contribution of the research work is given as follows:

- The RA scheme based on DRL is proposed in this research which enhances the D2D communication system's performance to a certain extent by dividing the vital sources such as spectrum, power and time over several devices. The interference issues in the D2D network are reduced while enhancing the throughput, as it enables the effective sharing of spectrum resources between the D2D pairs and cellular users. The efficient use of the available spectrum resources is possible by the suggested model to ensure reliable communication within the network.
- The A-DRL model is recommended in this work for the purpose of RA in the D2D cellular communication network. The A-DRL makes effective use of resources within the network to enhance the battery life while reducing interference issues. The throughputs of all cellular links are effectively maximized by the A-DRL since it learns to dynamically allocate the resources to the D2D links of the network. In addition, the A-DRL model considers the condition of the network for selecting the optimal resources to execute the efficient RA. Even in a high-density network, the A-DRL reliably performs the RA to reduce the latency while improving the throughput of the network.

The remaining parts of the paper are listed here: Relevant research work on the RA is given in Part 2. The description of RA in D2D networks based on RL is described in Part 3. The performance evaluation is given in Part 4. The summary of the research work is given in Part 5.

2. Works on Literature

The experts in the field of communication have performed research for handling the interference issue in D2D communication. The research work [1] proposed an improved Monte Carlo tree search for establishing the RA in the D2D network. Here, the parallel convolution network updates the parameters for solving the issues of transmit power. The dynamic RA output is attained as the result of the model, and it greatly reduces the outage probability within the D2D network. Article [2] proposed a combination of DRL-KM algorithm for effectively performing the spectrum allocation within the network. This model reduces the centralized computational overhead of the network while enhancing the convergence speed. Further, the signal scaling is reduced by utilizing the cooperative link set, which reduces the overhead issues in the D2D communication network. The HPSCAV was proposed by Pasha and Mohamad in [3] for efficiently allocating the resources in the D2D communication with minimum power. This model satisfies the QoS of the user, which ensures the energy efficiency of the network. In Chennaboin and Nandakumar in [4], the "HFDCSA" was proposed for minimizing the link rates in the D2D network. After that, energy efficiency and the sum rates are considered in the optimal relay selection to get high performance in the D2D communication network. Pan and Jieyang [5] proposed a DQN for performing the optimal RA within the network. The efficiency of the D2D system in the heterogeneous cellular network is effectively increased by this model. In addition, the QoS and stability of the network are also ensured using this approach.

3. Description of RA in D2D Communication Networks Based on RL

3.1 System Model of D2D Communication Networks

The system model of the D2D communication network is given in Fig. 1. The centre of the network consists of eNodeB [3]. Around the eNodeB of the D2D communication network, the DUE and CUE are located. When the channel is free, the CUE resource block is shared by the DUE. If the predetermined threshold is greater than the total interference of the CUD, then the CUE resources are shared by the D2D pairs in the D2D communication network. The term $\Re$ and $\Im$ represent the group of D2D pairs and channels within the network where $|\Re| = R$ and $\Im = J$, $\mathrm{M} = M$. The transmitter power of the jth D2D pair is represented as qp_j^o and the CUE's transmit power is denoted as qp_D^o; here the same channel o is shared by both D2D pair and CUE. The term CX specifies the bandwidth and spectral density is specified by the term OT_0. In between the kth D2D receiver and j^{th} D2D transmitter, the channel gain is $i_{j,k}^o$. In between the eNodeB and j^{th} D2D transmitter, the channel gain is expressed as $i_{j,0}^o$. The expression for the data rate of DUE Es_j is given in Eq. (1).

$$Es_j\left(\overrightarrow{qo}\right) = \sum_{m\in M} CX\log_2\left(1+\frac{i_{j.j}^o qp_j^o}{OT_0CX+\sum_{N\in\Re} i_{m.j}^o qp_m^o + i_{0,j}^o qp_D^o}\right) \quad (1)$$

Here, the term $\overrightarrow{qo} = \left\{qp_1^1, qp_1^2, \ldots\ldots qp_{\Re}^O,\right\}$. This research constructs the ADRL for the potential of optimizing the communication efficiency in the D2D communication network. The system model of the D2D cellular network is given in Fig. 1.

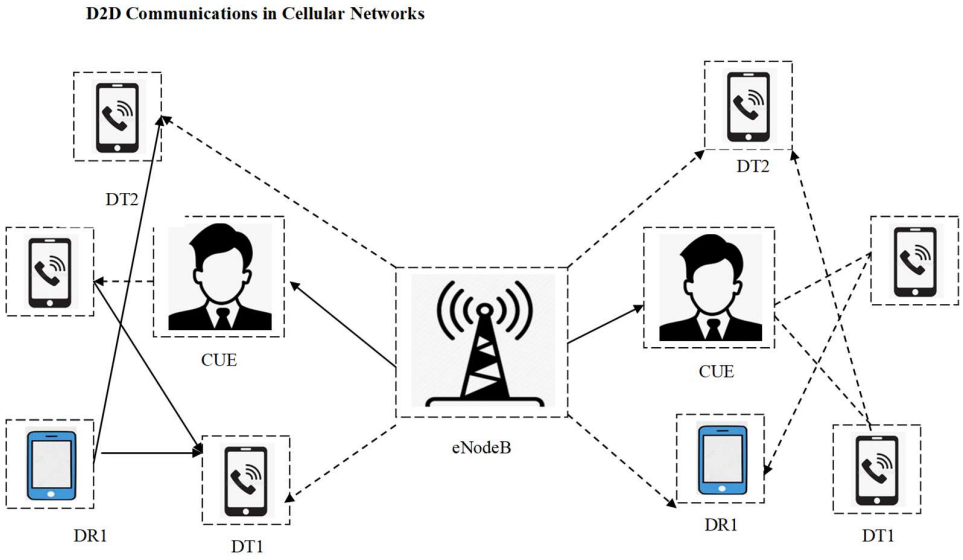

Fig. 1 System model of the D2D communication in cellular network

3.2 Problem Formulation

The optimization problem is formulated after modelling the D2D system. The recommended strategy maintains the communication quality and maximizes energy efficiency. The communication distance between the user and the transmit power is optimized to achieve this goal. The DUE data rate is greatly increased as a result of the effective RA in the D2D network. In addition, the DUE transmission interference is reduced below J but the data rate of the DUE is maintained to be no less than the Es_j . The formulation of the optimization problem is given in Eq. (2).

$$\begin{aligned} &\underset{0<(qo)}{Max} \sum_{j\in\Re} ES_j\left(\overrightarrow{pq}\right) \\ &subjected\ to \sum_{j\in\Im} qp_j^o \le qp_{\max} \quad \forall j \in \Re \\ &\sum_{j\in\Im} i_{j,0}^o pp_j^o \le J_{tnr} \quad \forall n \in \Im \\ &ES_{thr} \le ES_j\left(\overrightarrow{pq}\right) \quad \forall j \in \Re \end{aligned} \tag{2}$$

The maximum communication power of the DUE is denoted by the term $qp_{\max}$, the minimization of interference with the CUE is denoted as J_{tnr} and the constraint $\overrightarrow{pq}$ expresses the minimum data rate of the DUE. In the short computation time, the determination of the optimal solution is difficult for non-convex optimization because of the large D2D pairs. Therefore, this work developed A-DRL for performing the RA. Within a short period of time, the near-optimal solution is attained by this model. The QoS constraints $\sum_{j\in\Im} i_{j,0}^o pp_j^o \le J_{tnr}$ *and* $ES_{thr} \le ES_j\left(\overrightarrow{pq}\right)$ are violated and need to be satisfied under dense network conditions.

3.3 RL and its Preliminaries

In a dynamic and complex environment, effective RA is possible by the RL. The Markov decision process is designed as a joint optimization issue and it is approximated as $\tau 1$ [21]. Additionally, the Thompson Sampling method is used for improving the balance between exploration and exploitation in uncertain environments, which makes the RA more effective [22]. At the discrete time scale, the joint optimization problem is represented in the form of $\left(T, B, q, s^u\right)$. Here, the term T denotes the state space and the set of action space is denoted as B . From the present network state $t^u \in T$ to the new state $t^{u+1} \in T$, the state evolution probability is represented as t^u . At the time slot u , the immediate reward attained by the agent

$b^u \in B$ $ie, q\left(t^{u+1} \mid t^u, b^u\right)$ is denoted as s^u. From the state to action, the policy π performs the mapping process. The environment is not fully observed by the agents (D2D users). The agents (D2D users) optimize the strategies by interacting with the environment. At the time slot u, the network state of the agent j is expressed as y_j^u. According to the policy π_j and current state t_j^u, the action b_j^u is taken by j. Based on this action, the transmission channel and appropriate communication models are selected. The instant reward s_j^u, next state t_j^{u+1} and the next observation information are p_j^{u+1} selected after all decisions made by the agents. Until reaching the maximum episode, the above process is repeated by j. Thus, RL is constructed based on reward, action, observation and state.

3.4 A-DRL for RA

The A-DRL is used for RA in D2D cellular communication. The A-DRL model combines the strength of the RL and the attention mechanism of the DL model for executing the RA in the complex cellular network. Here, the most suitable server for the user is selected by incorporating the attention mechanism. The incorporation of the attention mechanism with DRL makes this model practically more efficient for RA. The ADRL model reduces the computational complexity and mainly aims to attain good performance in the D2D network. The major resources of the D2D network are D2D links and transmit power. The A-DRL-based RA enhances the throughput of the D2D links and cellular users over numerous time steps. The spectrum and the transit power of the devices in the D2D links and cellular system are obtained considering outage probabilities. The optimization problem required for RA is also modelled by the proposed model. The suggested ADRL introduced in this research work is well-suited for RA since it can easily interact with dynamic and unconstrained environments to reveal hidden patterns in the network structure. The signaling overhead issues in D2D communication are effectively solved using this model. Importantly, the ADRL does not require exact information on the system dynamics, which is very useful for maximizing long-term successful transmission in D2D communication. The incorporation of the attention mechanism assists the DRL in handling the large action space and it also performs the RA on the large-scale network. In the proposed A-DRL, agents are located on the central coordinator for executing the learning process. During RA, the proposed A-DRL reduces the computational overhead involved in the training and testing process. Based on the location of the participating devices, the resources are promptly allocated by the A-DRL. The attention function computes an output as the sum of input value vectors that are weighted by the input's compatibility. Here, the input key vectors for each value are derived from the query vector. The expression for the attention is given in the Eq. (3).

$$Att(Q,W,E) = STMX\left(\frac{QW^T}{\sqrt{t_W}}\right)E \tag{3}$$

Here, the query vector is represented as Q, and the value and key vector are denoted as E, W. The key vector dimension is denoted as t_W. The results of the attention from the multiple heads are determined using Eq. (4).

$$I = Att\left(QO_j^Q WO_j^W, EO_j^E\right) \tag{4}$$

$$MHA(Q,W,E) = cncate\left(G_1,....G_i\right)O^P \tag{5}$$

Here, the trainable weights are represented as O and the number of heads is denoted as i.

Initially, the resources of all D2D links are identified simultaneously. The single agent of the A-DRL is considered as the central coordinator of eNB and it is responsible for allocating the resources to all DUEs. The time duration for data transmission between the DUER and DUET is defined using an episode. Within each episode, the time steps are counted. Between 0 and $U-1$, the time step u is defined. The location of all UE in the cell is defined using the state t_u at time step u. The vector location of all DUER, DUET, and CUE is represented as a_u. At time step u, the allocated channel indices and transmit power of all D2D links are denoted as d_u and q_u, respectively. At the time step u, the channel indices and transmit power of all D2D links are used for defining the action b_u. The term s_u denotes the instantaneous reward at the time step u. It is also measured by averaging the cellular links and throughput of the D2D.

$$s_u = \frac{1}{L}\sum_{L=0}^{l-1} S_d\left(1-\gamma_{C,u}^{l}\right)+\frac{1}{N}\sum_{n=0}^{N-1}\sum_{l=0}^{L-1} \mathrm{H}_{nl,u} S_e\left(1-\tau_{n,u}^{l}\right) \quad (6)$$

The term $\widetilde{s}_u$ denotes accumulated reward up to $u-1$.

$$\widetilde{s}_u = \sum_{j=0}^{u-1} s_j \quad u \le U \quad (7)$$

The accumulated reward in terms of benefit is represented as benefit, and it is calculated using Eq. (8).

$$\widetilde{s}_U = \sum_{u=0}^{U-1} s_j \quad (8)$$

The benefits of D2D communication are enhanced as a result of the RA. The structural illustrations of the RA in the D2D communication network are given in Fig. 2.

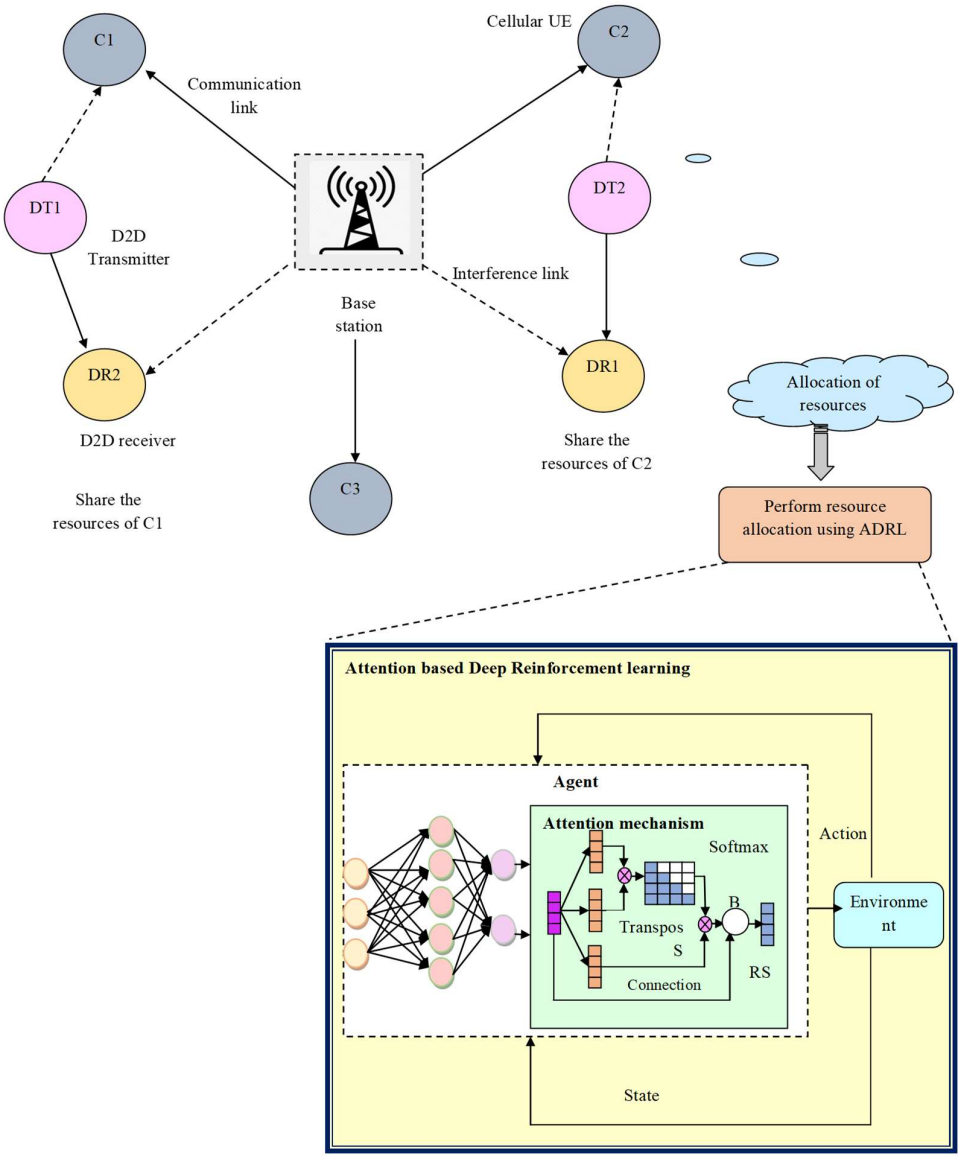

Fig. 2 Structural illustration of the RA in a D2D communication network

4. Performance Evaluation

The evaluation of the designed RA is discussed in this section. The recommended RA approach was executed using MATLAB2020a on the Windows 11 PC to showcase the efficiency of the model. Prior approaches such as DRL-KM [2], HFDCSA [4], DQN [5], and HPSCAV [3] were

employed to prove the efficiency of the designed model. The suggested methodology was validated by considering the following performance metrices: data rate, energy efficiency, execution time, Outage probability, SNR, and spectrum efficiency metrics.

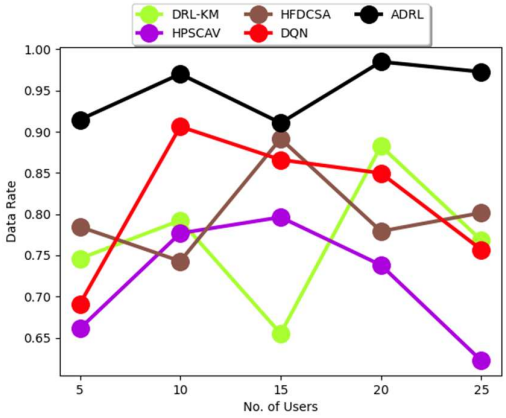

Fig. 3 Data rate

Fig. 7 evaluates the performance of the developed A-DRL-based RA schemes in terms of data rate and compares the performance with the other models such as DRL-KM, HFDCSA, DQN, and HPSCAV. As noted in Fig. 3, the proposed A-DRL model relatively attained a higher value of data rate as compared to the previous models such as DRL-KM, HFDCSA, DQN, and HPSCAV. However, the value of the data rate slightly decreased at user 15, and the same value of the data rate is attained at user 5. The advantage of the proposed A-DRL is evident from these results since it attained the maximum data rate at the 20th user. The improved capability of data rate showcases the enhanced communication capability of the designed A-DRL.

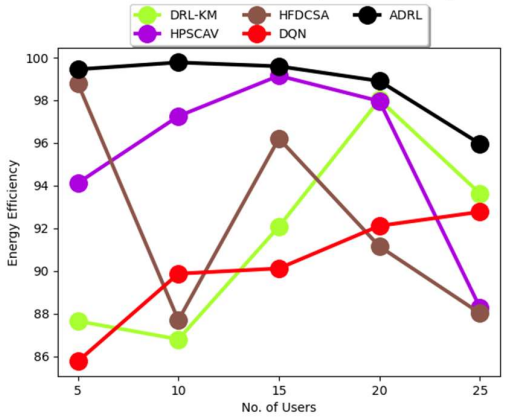

Fig. 4 Energy efficiency

The energy efficiency trend of the proposed A-DRL is given in Fig. 4. It can be seen from Fig. 4 that the recommended A-DRL's energy efficiency decreases gradually from user 5 to user 25. However, the proposed A-DRL successfully attained higher energy efficiency over the other models like DRL-KM, HFDCSA, DQN, and HPSCAV. At the same time, the DRL-KM model reaches a slightly lower value of energy efficiency at the 25th user, which proves the efficiency of the recommended model in the RA. At the 15th user, the recommended A-DRL and HPSCAV nearly provide the same energy efficiency during the RA. Thus, the results indicate that the recommended A-DRL effectively handles the challenge of increasing data traffic in the communication process.

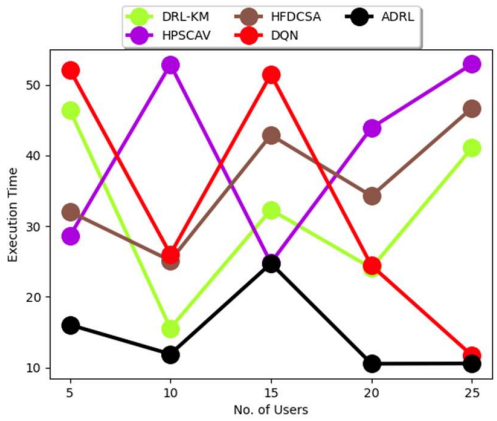

Fig. 5 Execution time

The execution time of the proposed A-DRL is revealed in Fig. 5. The in-depth evaluation of the execution time of the recommended A-DRL allows us to comprehensively determine the effects of the recommended model in RA. Here, the suggested A-DRL reaches a lower execution time while executing RA. It is noted from Fig. 5 that the suggested A-DRL has direct effects on improving the communication of the D2D network. In contrast, the conventional DRL-KM, HFDCSA, DQN, and HPSCAV consume higher execution time during RA. However, the proposed A-DRL consumes fewer resources and effectively allocates the resources within minimum time, improving the communication performance of the D2D network.

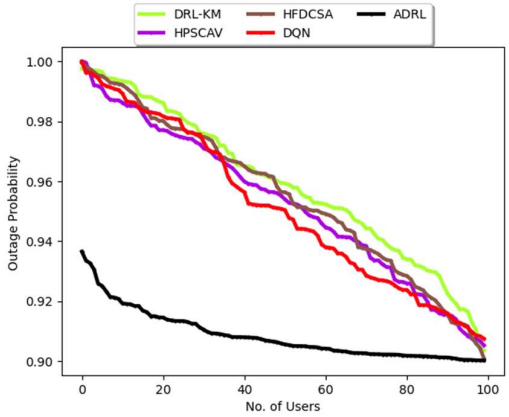

Fig. 6 Outage probability

Fig. 6 reflects the outage probability of the recommended A-DRL. When the user count is improved from 0 to 100, the outage probability of the suggested A-DRL is gradually decreased. Whereas the conventional models like DRL-KM, HFDCSA, DQN, and HPSCAV exhibit higher outage probabilities. The results show that the suggested A-DRL attained higher performance gain, which means it can efficiently manage network performance even in disturbed environments, as it effectively utilizes the necessary resources to optimize the performance of the D2D cellular communication network.

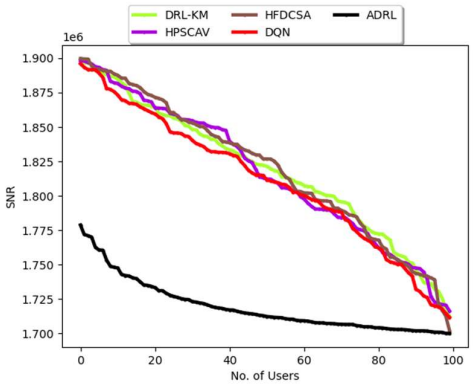

Fig.7 SNR

The SNR of the proposed A-DRL is noticeably decreased in Fig. 7 at the 100th user. Here, the SNR of all models is considerably decreased with an increase in the numbers of users. However, notably, the proposed A-DRL model still shows lower SNR of 1.7 at the 100th user. With the lower SNR value, the adaptability of the proposed ADRL is superior to other models such as DRL-KM, HFDCSA, DQN, and HPSCAV. The lower SNR of the designed A-DRL reflects its stronger performance and anti-interference capability. The recommended A-DRL model helps make effective decisions for RA.

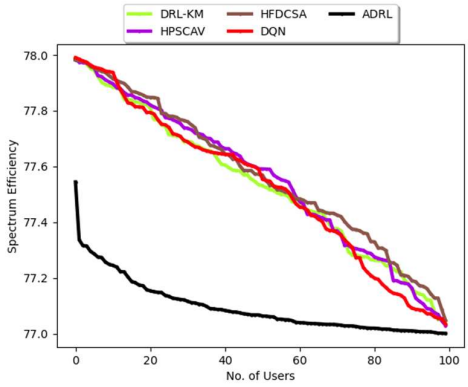

Fig 8. Spectrum efficiency

The visual comparison of the proposed A-DRL's spectrum efficiency against other models is given in Fig. 8. It can be noted from Fig. 8 that the spectrum efficiency of the proposed A-DRL is lower than that of DRL-KM, HFDCSA, DQN, and HPSCAV. This is mainly because the proposed model's learning capability is more sophisticated than that of the DRL models due to the incorporation of the attention module. Here, each agent of the proposed A-DRL is trained individually to understand how subsequent agents act in the RA process. In addition, the designed A-DRL considers the user's location and the target rate of the cellular link to allocate the communication resources within the D2D network.

5. Conclusion

In this research, an efficient RA scheme was proposed based on DRL for eliminating interference issues and enhancing the D2D communication network's performance. Energy-efficient communication in the complex D2D cellular network was attained through the proposed model. The communication power of the D2D communication network was adjusted via the recommended model. The A-DRL was the central contribution of the research work, which was more effective for collaborative and intelligent decision-making over the D2D user in RA, and significantly boosted the performance of the D2D network. This research will be extended in the future to perform RA under multi-cell scenarios in a multi-D2D collaborative network.

References

[1] XINZHOU LI and GUIFENCHEN, "D2D Cooperative Communication Network Resource Allocation Algorithm Based on Improved Monte Carlo Tree Search," IEEE Access, 2023.

[2] YANG YU, AND XIAOQING TANG, "Hybrid Centralized-Distributed Resource Allocation Based on Deep Reinforcement Learning for Cooperative D2D Communications," IEEE Access, 2024.

[3] Shaik Ahmed Pasha and Noor Mohammed Vali Mohamad, "A modified LSTM with QoS aware hybrid AVO algorithm to enhance resource allocation in D2D communication," Pasha and Mohamad J Wireless Com Network, vol. 12, 2024.

[4] Ramesh Babu Chennaboin1 and S. Nandakumar, " An adaptive MLP-based joint optimization of resource allocation and relay selection in device-to-device communication using hybrid meta-heuristic algorithm," Chennaboin and Nandakumar J Wireless Com Network, vol. 50, 2024.

[5] ZIYU PAN AND JIEYANG, "Deep Reinforcement Learning-Based Optimization Method for D2D Communication Energy Efficiency in Heterogeneous Cellular Networks," IEEE Access, 2024.

[6] Patil, Priyanka, and Vaibhav S. Hendre, "Modified archimedes optimization algorithm-based resources allocation in multicell D2D communications," Cluster Computing, vol. 28, no. 5, pp. 1-13, 2025.

[7] Wu, Xiaobin, A. Vijayaraj, Nadia Sarhan, K. Chanthirasekaran, Emad Mahrous Awwad, and Guma Ali, "An Effective Resource Allocation Framework for D2D Communication in 5G Wireless Personal Area Network With Hybrid Optimizer," International Journal of Communication Systems, vol. 38, no. 4, pp. e6092, 2025.

[8] Dobruna, Jeta, Zana Limani Fazliu, Iztok Humar, and Mojca Volk, "On Optimizing Resource Allocation: A Comparative Review of Resource Allocation Strategies in HetNets," Computer Modeling in Engineering & Sciences (CMES), vol. 142, no. 3, 2025.

[9] Chandra, Saurabh, Rajeev Arya, and Maheshwari Prasad Singh, "Age of information-aware intelligent resource management in D2D-enabled social IoT networks," Computers and Electrical Engineering, vol. 123, pp. 110295, 2025.

[10] Jian, Zhaomin, Chao Ma, Yunchao Song, Mengshuang Liu, and Huibin Liang, "Flexible Resource Optimization for D2D XL-MIMO Communication via Adversarial Multi-Armed Bandit," Electronics, vol. 14, no. 8, pp. 1498, 2025.

[11] Jiang, Tao, Ming Jin, Qinghua Guo, Yinhong Liu, Yaming Li, and Junteng Yao, "Full-duplex ISAC-enabled D2D underlaid cellular networks: joint transceiver beamforming and power allocation," IEEE Transactions on Cognitive Communications and Networking, 2025.

[12] Tseng, Shu-Ming, Sz-Tze Wen, Chao Fang, and Mehdi Norouzi, "Cross Layer Power Allocation by Graph Neural Networks in Heterogeneous D2D Video Communications," IEEE Access, 2025.

[13] Pei, Errong, Xinhu Chen, Niexin Xiang, Chenkai Ren, and Yonggang Li, "Joint 3D Trajectory and Resource Optimization in Licensed and Unlicensed Spectrum for D2D-Underlaid UAV Communication Networks," IEEE Internet of Things Journal, 2025.

[14] Patel, Rachit, and Rajeev Arya, "Resource allocation for D2D underlay communication in 6G networks considering 3D distance path loss model," In Intelligent Computing and Communication Techniques, pp. 269-274, CRC Press, 2025.

[15] Ding, Hanxue, Shaoyi Xu, Ziheng Xu, Rongtao Xu, and Bo Ai, "Achieving Deterministic Delay in Industrial Internet of Things: Resource Allocation Schemes in the Finite Blocklength Regime," IEEE Transactions on Vehicular Technology, 2025.

[16] Shafaq, Tayyaba, Jawad Mirza, Waeal J. Obidallah, and Mohammed Alkhatham, "An energy-efficient task offloading in D2D-assisted IoT networks using matching algorithms," The Journal of Supercomputing, vol. 81, no. 4,pp. 568, 2025.

[17] Parashar, Sadhvi, and Rajeev Arya, "MA2CL: Multi-Agent Actor-Critic Learning Scheme for Efficient Resource Management in 5G-Enabled NB-IoT Networks," Internet Technology Letters, vol. 8, no. 3, pp. e70011, 2025.

[18] Liu, Peng, Zixu Zhang, Chenhao Ren, and Hailong You, "A blockchain-based resource sharing incentivization mechanism for multi-to-multi in compute first networking," Computer Networks, pp. 111318, 2025.

[19] Sesli, Erhan, Erdal Gürsoy, and Gökçe Hacıoglu, "Energy-efficient OFDMA resource allocation in HetNets using discrete particle swarm optimization," Peer-to-Peer Networking and Applications, vol. 18, no. 3, pp. 125, 2025.

[20] Bajpai, Rahul, Sukriti Mishra, and Naveen Gupta, "NOMA and Millimeter Wave Assisted SWIPT for Full-Duplex Cooperative D2D Communications," IEEE Access, 2025.

[21] Zhi, Yuan, Jie Tian, Xiaofang Deng, Jingping Qiao, and Dianjie Lu, "Deep reinforcement learning-based resource allocation for D2D communications in heterogeneous cellular networks," Digital Communications and Networks, vol. 8, no. 5, pp. 834-842, 2022.

[22] Russo, Enrico, Maurizio Palesi, Davide Patti, Giuseppe Ascia, and Vincenzo Catania, "Attention-based deep reinforcement learning for qubit allocation in modular quantum architectures," arXiv preprint arXiv:2406.11452, 2024.