International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(4s) 2026 pp. 1066-1076
https://doi.org/10.70917/ijcisim-2026-2619
Article

Deep Reinforcement Learning Framework for Efficient Resource Allocation in NOMA Systems

Shruthi N[1,*], K. Ramesha[2], Sumit Gupta[3]

[1,*]Assistant professor, Department of Electronics and Telecommunication Engineering,
Bangalore Institute of Technology, Bengaluru, India. Email: shruthin@bit-bangalore.edu.in
[2]Department of Electronics and Instrumentation Engineering, Dr. Ambedkar Institute of Technology, Bangalore, India
[3]Research Head, R&D Department, Xsys Software Technologies, Bangalore, India. Email: sumit@xsyssoftwaretechnologies.com
*Corresponding author - shruthin@bit-bangalore.edu.in

Abstract: The rapid growth of next-generation wireless networks has led to a substantial proliferation in the use of smart devices and emerging applications, which demand high computational power and low latency. Managing these devices within limited resources presents a substantial challenge. Non-Orthogonal Multiple Access (NOMA) has gained huge attention as a favourable technique to address these challenges by allowing multiple signals to be transmitted and received concurrently on the same frequency band, thereby improving spectral efficiency, network capacity, and energy efficiency. In this work, we focus on optimizing resource allocation in NOMA systems, particularly in the context of 5G networks. Despite the advantages of NOMA over traditional Orthogonal Multiple Access (OMA) methods, efficient dynamic resource allocation remains a critical challenge. To address this, we propose a novel model that leverages deep reinforcement learning for optimized resource allocation in NOMA systems. The proposed model demonstrates significant improvements in energy efficiency and spectral efficiency compared to existing methods, while also addressing challenges related to power allocation, user clustering, and interference management. Our approach provides a scalable and adaptive solution for enhancing the performance of NOMA-enabled 5G networks.
Keywords: NOMA, Deep Reinforcemenr Learning, Resource allocation, Energy efficiency, Spectral efficiency

1. INTRODUCTION

Recently, we have noticed a tremendous growth in wireless and mobile communication domain. This rapid growth of next-generation wireless networks has led to an unparalleled surge in the use of smart devices, such as tablets, smartphones, and wearables, along with the emergence of new applications [1]. The widespread deployment of these devices presents the challenge of managing them within limited resources. Furthermore, many of these emerging applications demand significant computational power and low latency, making it difficult for power-limited and size-constrained devices to provide the desired quality of service under such conditions [2]. Currently, the NOMA has emerged as promising technique to handle these challenges. It is a communication technique that allows multiple signals to be transmitted and received concurrently on the same frequency band [3]. Orthogonal multiple access (OMA) is deliberated one of the promising technique in this domain however, the NOMA offers several advantages over traditional methods such as increased spectrum efficiency, handling more number of devices to expand the network capacity, minimizing the latency and singnalling cost [4, 5]. In OMA, each user is assigned a separate orthogonal channel, while in NOMA, users are allocated different power levels and superposition coding is utilized. This enables the decoding of signals through successive interference cancellation (SIC) based on their respective power levels [6]. The benefits of NOMA over OMA include higher spectral efficiency, expanded network capacity, and better energy efficiency. Additionally, NOMA enhances energy efficiency by enabling resource sharing among users, thereby decreasing the energy required for communication. Generally, the NOMA systems can be categorized into two main categories as power-domain and code-domain NOMA systems.

Power domain NOMA: In this approach, users are categorized based on their power levels. The core idea is that users experiencing better channel conditions receive lower power levels, whereas those with poorer channel conditions are assigned higher power levels [7]. Superposition coding is used to combine the signals of multiple users into one transmission. The receiver end uses SIC to decode the obtained signal. This approach improves spectrum efficiency and network capacity because it allows numerous users to share the same band simultaneously

Code domain NOMA: in this, users are identified by unique spreading codes rather than power levels. This technique is similar to the traditional CDMA but differs in that it allows non-orthogonal codes to be used, which means the codes are not necessarily independent of each other. The receiver uses advanced detection techniques, such as message passing or maximum likelihood detection, to separate the signals from different users. Code Domain NOMA can support massive connectivity even with overlapping codes [8]. It also provides flexibility in resource allocation and can be combined with Power Domain NOMA to further enhance system performance.

Generally, the combined signal is received at the receiver end which need to be separated by the multiuser detection mechanisms such as successive interference cancellation [9]. This process leads to demonstrate the resource allocation optimization problem in NOMA and other OMA systems. The decoding process in NOMA systems leads to a complex resource allocation optimization problem, especially as the number of users and devices grows [10]. To address this challenge, machine learning (ML), deep learning (DL), and deep reinforcement learning (DRL) have emerged as powerful tools for optimizing resource allocation in NOMA systems. These techniques allow for dynamic, data-driven decision-making, improving efficiency and performance in real-time. Deep learning models are particularly effective at analyzing large datasets and extracting patterns, making them ideal for tasks like predicting user behavior, channel conditions, and traffic patterns. By leveraging deep learning architectures, NOMA systems can intelligently allocate resources based on real-time data, optimizing performance metrics such as spectral efficiency and latency. Deep reinforcement learning takes this a step further by enabling systems to learn optimal resource allocation policies through interaction with the network environment. DRL can model the resource allocation problem as a Markov Decision Process (MDP), allowing the system to make adaptive, intelligent decisions based on network state transitions. Techniques like Deep Q-Networks (DQN), Proximal Policy Optimization (PPO), and Actor-Critic methods are particularly useful in this context, as they enable the system to learn from its environment and optimize resource allocation in both power-domain and code-domain NOMA systems. By combining deep learning and DRL, NOMA systems can overcome the limitations of traditional optimization methods, which often require predefined models of the environment. Instead, these advanced techniques allow the system to learn directly from interactions with the network, making them well-suited for the dynamic, real-time scenarios typical of next-generation wireless networks.

Thus, the integration of deep learning, machine learning, and DRL into NOMA resource allocation strategies offers a promising approach to overcoming the challenges posed by modern wireless networks. These intelligent, data-driven methods provide a pathway to achieving higher spectral efficiency, better energy efficiency, and improved overall quality of service in the ever-expanding ecosystem of smart devices and applications. Therefore, this work mainly focuses on development of deep reinforcement learning based approach for resource allocation in NOMA systems. The main contributions of this work include deep reinforcement learning based approach for optimal resource and power allocation. in order to achieve this, we presented a combined problem formulation for both tasks and presented action, state and reward mechanism for each phase.

The remainder of the manuscript is organized in following sections: section II presents the brief literature review where existing schemes of resource allocation are discussed, section III presents the proposed deep reinforcement learning based solution to enhance the resource allocation, section IV presents the outcome of proposed model and its comparative analysis with existing methods, and, finally, section V presents the concluding remarks about this research.

2. LITERATURE SURVEY

This section presents the brief literature review about existing methods of optimized resource allocation in NOMA system. Recently, several methods have been introduced to address the resource allocation related issues and to augment the spectral efficiency of the system. Fang et al. [11] focused on combining the Multi-access edge computing (MEC) and NOMA systems because MEC is capable to enhance the computing capacity whereas NOMA can be useful in providing high data rates. Therefore, this article explores delay suppression in NOMA-MEC networks. The authors adopt a partial offloading policy. The main aim of this model is to minimize the delay among users which is achieved by optimizing the task partition ratios.

Therefore, a delay minimization problem is formulated as a non-convex problem and later transformed into quasi-convex problem. Dong et al. [12] discussed that small cell networks are deployed densely to meet the criteria of next generation mobile communication. Generally, these small cell networks utilize backhaul links to produce multi-hop topologies, however, these links fail to control the unstable traffic. Therefore, authors have introduced adaptive topology based links to introduce multi-hop topologies. This system helps to minimize the transmission cost. Further, a graph theory based model is developed to dynamically manage the backhaul architecture. Shan et al. [23] reported the performance related issues in VANET communication systems and reported that the performance of these communication systems can be enhanced by incorporating NOMA systems. However, the performance of these NOMA systems is affected due to interference due to V2X communications. Therefore, authors introduced energy efficient resource allocation mechanism in V2X networks with the help of NOMA systems. According to this approach, a two-layer block coordinate descent model is developed to solve the power allocation sub-problem by introducing intra-group power allocation strategy in inner layer followed by Dinkelbach and concave–convex procedure to consider the lower bound of the problem. Further, spectrum sharing sub-problem is addressed with the help of energy efficiency maximization mechanism.

Chandra et al. [14] reported the importance of MIMO with NOMA in improving the overall performance of the 5G wireless communication system however this mechanism leads to increase the number of antenna arrays as the number of user increases. Beamforming techniques have been used widely to address these issues. In this work, a combined MMSE with zero forcing precoding model is presented to overcome the issues of inter-cluster interference. The beamforming process is performed on the user clustering performed by applying fuzzy k means approach. Later, hybrid artificial gorilla troop with leader optimization is applied for optimal resource and power allocation optimization.

Guo et al. [16] addressed the challenge of optimizing energy efficiency in a CNOMA network. This problem is tackled in two phases: first, by deriving optimal closed-form solutions for power control, and second, by framing the system’s EE optimization as a self-play Go game, where the goal is to achieve the highest EE. To solve this, a deep Monte Carlo tree search model is introduced, which is directed by a neural network to explore various possible trajectories and assess the resulting EE rewards. The neural network is trained to predict move choices and determine game outcomes.

Kim et al. [21] developed a sub-channel and power allocation for MC-NOMA systems which plays significant role in IoT based systems by allowing multiple sub-channels to facilitate the massive connectivity. In order to obtain the solution, this article formulates a binary decision problem and introduced a deep learning based solution to enhance the maximum sub-channel power for NOMA users. Moreover, a user selection algorithm is also introduced to avoid the performance loss caused due to outage users.

Beuria et al. [22] focused on cooperative NOMA approach to improve the spectral efficiency of the communication system. The existing methods have reported that the existing researches have emphasized on optimal relay section and incorporated deep learning based model for power allocation tasks. The optimal relay selection plays crucial role to facilitate the data transmission to far users where this articles employs Maximum Energy Harvested Relay selection mechanism which selects the relay nodes based on their corresponding energy harvesting capabilities. Later, a deep convolutional optimized neural network to power allocation tasks. Further, the Grey Wolf Optimization scheme is also incorporated to fine-tune the weights of the deep learning network. Louis et al. [24] focused on energy and spectrum efficiency by designing a user grouping method which is designed to connect the near-far and far-users with C3. Moreover, this approach also uses a hybrid optimization method by combining coati and bald eagle optimization algorithms. Ghanbarzadeh et al. [25] focused on channel and power allocation in NOMA systems by combining convex optimization and Machine Learning approach. This work proposed an ensemble model where Convolutional Neural Network, Neural network and Random Forest classifiers are used to formulate the ensemble classification model. Shyamala et al. [26] presented a game theoretic approch for delay tolerant networks with cost and congestion optimization.

2.1. Research gap and challenges

The discussion presented in previous section describes the existing method to address the resource allocation related issues NOMA systems for advanced cellular communication systems. Despite of significant advancements, these systems face several challenges. In the context of 5G systems, NOMA combined with resource allocation presents a significant area of research, particularly in addressing the challenges of optimizing spectral efficiency, energy efficiency, and system capacity. A key research gap lies in the development of efficient algorithms for dynamic resource allocation in NOMA-enabled 5G networks. These algorithms must effectively manage the complex interplay between power allocation, user clustering, and interference management, while ensuring fairness among users with diverse QoS requirements. Moreover, the integration of NOMA with emerging 5G technologies such as massive MIMO, millimeter-wave communications, and edge computing introduces additional layers of complexity. The dynamic nature of user mobility and the heterogeneous network environment of 5G further complicate resource allocation, requiring adaptive and scalable solutions. Furthermore, security and privacy concerns in NOMA-based 5G networks, particularly in the context of resource allocation, remain underexplored, presenting a critical challenge as 5G systems aim to support a vast number of devices and applications.

The rapid expansion of wireless and mobile communication technologies has driven an unprecedented surge in the use of smart devices and other IoT devices. This proliferation is accompanied by the emergence of demanding applications like augmented reality (AR), autonomous driving, and tele-surgery, all of which require substantial computational power and minimal latency. Consequently, managing these devices and their communication needs within the constraints of limited spectrum and computational resources presents a significant challenge.

Traditional OMA techniques, while effective, face limitations in addressing the increasing demand for higher data rates and improved network efficiency. OMA methods allocate distinct frequency channels or time slots to individual users, which can lead to inefficiencies in spectrum utilization and increased latency due to the separation of users' signals. Despite its advantages, implementing NOMA effectively involves several challenges:

- Spectrum Efficiency: Efficiently managing spectrum resources in a way that maximizes throughput while minimizing interference.

- User Pairing and Resource Allocation: Optimal user pairing and resource allocation strategies are needed to balance performance and resource utilization.

- Energy Efficiency: Reducing energy consumption while maintaining or improving communication performance remains a key concern, especially in power-limited devices.

- Computational Complexity: Managing the computational load associated with advanced NOMA techniques, including deep reinforcement learning for optimization, within constrained devices.

This problem necessitates the development of novel strategies and models to leverage NOMA's benefits fully. It involves exploring advanced techniques such as deep reinforcement learning to optimize user pairing, resource allocation, and energy efficiency, ultimately addressing the demands of next-generation wireless networks while overcoming the limitations of traditional OMA methods.

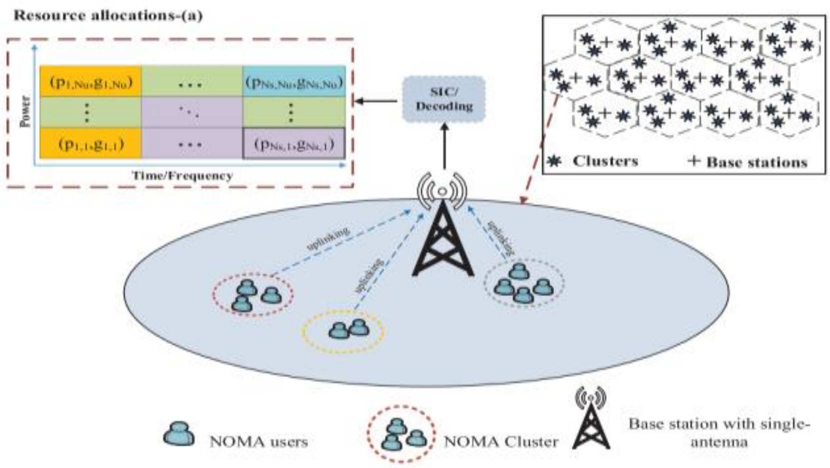

Figure 1 NOMA network model

3. METHOD

This section presents the detailed proposed approach for resource allocation in NOMA enabled 5G communication system. The first subsection presents the overview of system model, next subsection presents NOMA user and signal model, finally, the proposed deep reinforcement learning approach is presented to derive the final allocation procedure

3.1. System Model

Let us consider that the below given figure 1 represents an upsampling network enabled with NOMA technique where N_b represents the base station communicate with $N_u(t)$ with the help of N_s orthogonal sub-channels. The base station and sub-channel are indexed as $\phi_b = \{b_1, \dots b_{N_b}\}$ and $\phi_s = \{s_1, \dots s_N\}$ for base station and sub-channel respectively. In this model, we consider that a set of user is served by a corresponding base station as $b_i \in \phi_b (i \in [1, N_b])$ with the help of sub-channel given as $s_j \in \phi_s$ $(j \in [1, N_s])$, thus, this can be represented as $\phi_u^{i,j} = \{u_i, \dots, u_{N^{i-j}}\}$ where $N_u^{i,j}$ represents the number of the intra-set uses. In this model, the base station and users are considered to be equipped with the single antenna and bandwidth for each user is equally divided into N_s sub-channels resulting in $\frac{B}{N_s}$ as the bandwidth for each sub-channel.

3.2. NOMA user and signal model

According to the NOMA principle, the same resource block can be used to serve more number of users. These users which are use same resources can be grouped together. In this work, each sub-channel represents one NOMA cluster where $N_u^{i,j} \geq 2$. Here, we assume that the base stations contain perfect CSI of all users. Based on this type of CSI information, the base stations are can optimize the sub-channel allocation for the active users. For any user u_k, the grouping variable at time t can be defined as:

$$g_k^{i,j}(t) = \{1, u_k \; connects \; to \; BS \; with \; the \; help \; of \; subchannel \; s_j \; 0, otherwise$$

For this model, the NOMA group, the base station receives the superposed message from the available active users in $\phi_u^{i,j}$ and then performs SIC to decode the user's signal. In a considered time slot t, the SINR for users which are present the group can be represented as:

$$\gamma_k^{i,j}(t) = \frac{c_k^{i,j}(t) p_k^{i,j}(t) g_k^{i,j}(t)}{\sum_{k'=1}^{k-1} \quad c_k^{i,j}(t) p_k^{i,j}(t) g_k^{i,j}(t) + I_{inter}(t) + \sigma^2(t)}$$

Where,

$$I_{inter}(t) = \sum_{i' \in \phi_b} \quad \sum_{k' \in \phi_u^{i',j}} \quad c_{k'}^{i',j}(t) p_{k'}^{i,j}(t) g_k^{i',j}(t)$$

Similarly, the data rate for the user can be expressed as:

$$R_k^{i,j}(t) = \frac{B}{N_s} log_2 \left(1 + \gamma_k^{i,j}(t)\right)$$

3.3. Problem formulation, action, state and reward function

In order to obtain the optimized resource management, we mainly focus on subcarrier assignment and efficient allocation of power to enhance the overall weighted-sum throughput. Based on this, the joint resource management problem can be expressed as:

$$\sum_{m=1}^{M} \quad \alpha_m \left(\sum_{i=1}^{N_F} \quad \sum_{j=1}^{J_i} \quad v_{i,j}^m R_{i,j} \right)$$

Where $0 < \alpha_m < 1$ represents the weight of user m which is set as $\alpha_m = \frac{d_m}{(d_i)}$ where represents the distance from user m to BS.

This problem is further divided into two sub-tasks as user grouping and power allocation. During this process, the user grouping tasks is allocated first and then based on the output of this, the second sub-task of power allocation is executed. In order to solve this problem, the Reinforcement Learning is considered as promising technique. Generally, the conventional RL mechanisms are defined as Markov decision process. According to this mechanism, at each decision stage t, the assigned agent observers the current state $s_t \in S$ and executes the corresponding action $a_t \in A$ and receives the reward r_t for this action. With the help of these actions from the pre-defined policy π, the agents establish the continuous interaction with the environment E to maximize the future rewards.

(a) Action

This subsection presents the action modeling for the aforementioned two sub-tasks. The action a^u is given as $a_t^u = \left[\pi_{1,t}^u, .., \pi_{N_F}^u, t\right]^T, \pi_{i,t}^u \in \{0,1\}, \forall_i$ here, if any subcarrier i is assigned to the user of t^{th} step then $\pi_{i,t}^u = 1$ otherwise it is 0. Similarly, the action agent for power allocation task are given as $a_{m,t}^p = \left[\phi_{1,m,t}, \dots, \phi_{i,m,t}, \dots, \phi_{N_F,m,t}\right]^T$ and $\phi_{i,m,t} \in \{-1,0,1\}, \forall_i$. the value of $\phi_{i,m,t} = 1$ represents that the power allocation should increase and $\phi_{i,m,t} = -1$ represents that the power should decrease and $\phi_{i,m,t} = 0$ represents the power allocation remains unchanged. Here, the magnitude of power change is expressed by v and $v_{m,t}$ is the power indictor expressed as $v_{m,t} = \left[v_{1,m,t}, \dots, v_{N_F,m,t}\right]$. In this phase, the actual allocated power of user m on a given subcarried i is expressed as:

$$p_{i,m,t} = P_{total} \frac{v_{i,m,t}}{\sum_{m=1}^{M} \sum_{i=1}^{N_F} v_{i,m,t}}$$

(b) State

This section presents the state analysis mechanism for the aforementioned action. According to this approach, the user grouping subtask considers user priorities, QoS constraints, channel gain, occupied channels which is expressed as: $s_t^u = \{W, R^{mi}, H, O_t\}$ where W, R^{mi}, H, O_t parameters whether the user m is assigned to the subcarried i in the given step t, if user is assigned to the subcarrier then $o_{i,m,t} = 1$ otherwise $o_{i,m,t} = 0$. On the other hand, the power allocation subtask the M users are M agents and for each agent the corresponding state $s_{m,t}^p$ includes several information such as priority of users, QoS factor, channel gain and outcome of user grouping stage. Thus, it contains own information and other agent's information as $s_{m,t}^p = \{s_{self,m,t}^p, s_{other,m,t}^p\}$

(c) Designing the reward function

As discussed before, the complete mechanism is divided into two sub-tasks where first task considers the reward computation for the user grouping and the second subtask is assigned for the power allocation. The reward function for user grouping can be expressed as:

$$r_t^u = w^u \Gamma\left(\sum_{m=1}^{M} \sum_{j=1}^{J_i} v_{i,j}^m \leq N_{max}\right)$$

Where, w^u is the penalty coefficient. Similarly, the reward function for power allocation sub-task is expressed as:

$$r_{m,t}^{p,int} = w_I^p \Gamma\left(\sum_{m=1}^{M} \sum_{j=1}^{J_i} v_{i,j}^m \leq N_{max}\right) + w_{II}^p \Gamma\left(\sum_{i=1}^{N_F} R_{i,m} - R_m^{min}\right)$$

3.4. *Deep reinforcement learning resource allocation*

Various methods have been developed based on Q learning based mechanism to optimize the resource allocation performance but traditional Q learning methods require high memory space and also suffer from complexity issues. Moreover, these methods suffer from the poor learning capability due to complex data patterns. In order to overcome the complexity and memory, we adopted deep learning mechanism and introduced a novel deep reinforcement learning model as $Q(s, a: \theta)$ to produce the Q values with the help of θ. Thefore, the deep learning baed Q learning model mainly focus on memorizing θ instead of storing all states, their rewards and corresponding action pairs. The process of Deep learning based Q learning is demonstrated below:

(a) Initially, a state space S is considered which is unique, given as input to the DNN. At this stage, each state is formed by combining multiple sub-sets users, base station and sub-channels. Moreover, it also consists of current and previous rewards of the system
(b) R represents the reward of the model which is expressed by $R = \{r_i, r_l\}$ where r_i is the instantaneous reward and r_l is the long term average reward for the time slot t
(c) The actions for these schemes are represented in the form of multi-dimensional matrix represented as $A = \{a_1, a_2, .., a_n\}$. The deep reinforcement learning approach uses a loss function to compute θ based on its historical experiences which can be expressed as:

$$loss(\theta) = \frac{1}{N_e}\sum_{t=1}^{N_e} [y_t^{DRL} - Q(s_t, a_t; \theta)]^2$$

Where $y_t^{DRL} = r + \gamma Q(s', a'; \theta')$

In this approach, the gradient decent mechanism is applied to reduce the prediction error with the help of loss function. The θ is updated by analysing the new experience, which can be expressed as:

$$\theta \leftarrow \theta - [y^{DRL} - Q(s, a; \theta)]\nabla Q(s, a; \theta)$$

$$q_\pi(s, a) = r(s, a) + \gamma \sum_{s' \in S} \sum_{a' \in A} p_{ss'}(a) q_\pi(s', a')$$

$$q_{\pi^*}(s, a) = r(s, a) + \gamma \sum_{s' \in S} p_{ss'}(a) q_{\pi^*}(s', a')$$

Where $q_{\pi^*}(s, a)$ represents the Q values for DRL, with the help of discount factor the optimal reward policy can be expressed as:

$$\pi^*(s) = arg\ \underset{a \in A}{argmax}\ [q_{\pi^*}(s, a)]\ , \forall\ s \in S$$

Where $\pi^*(s)$ shows the optimal policy and optimal policy values for each state are obtained from this policy in any state s and action a. This can be expressed as:

$$Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha[r(s, a) + \gamma Q(s', a')\ ]$$

$Q(s, a)$ represents the updated Q values according the Bellman equation.

	Algorithm 1:
1.	Inputs: episodes, exploration trials, learning rate
2.	Initialization of DRL: network parameters, DL parameters
3.	Output: reward, and selection of best action to enhance the resource allocation
4.	Train DRL to obtain the good policy for resource allocation as θ
5.	For iteration =1: Ndo
6.	For iteration = 1: T do
7.	Compute reward $r(s_t, s_{t+2}, a_t)$
8.	Updated reward as $q_{\pi^*}(s, a) = r(s, a) + \gamma \sum_{s' \in S} p_{ss'}(a) q_{\pi^*}(s', a')$
9.	Compute loss as $loss(\theta) = [y_t^{DL} - Q(s_t, a_t; \theta)]^2$
10.	Update the prediction $y_t^{DL} = r + \gamma Q(s', a'; \theta)$
11.	Update mini-batches as historical experiences
12.	Return $Q(W_a)$

4. RESULTS AND DISCUSSION

This section presents the outcome of proposed deep reinforcement learning approach for resource allocation in advanced communication systems. The outcome of proposed model is compared with the existing state-of-art methods to demonstrate the robust performance of proposed model. The obtained performance is compared with the existing methods. Below given table 1 shows the complete details of simulation parameters used in this work.

Table 1 Simulation parameters

Parameter	Considered Values
Number of BS	4
Number of users	10
Max. power (macro-BS)	43 dBm
Max. power of small BS	40 dBm
Bandwidth	100 MHz
Path loss exponent	3.0

The outcome of proposed approach is compared in terms of average energy efficiency for varied number of users, throughput. Below given figure 2 illustrates the comparative analysis of the proposed approach with existing exhaustive search, Deep MCTS, DQN , CCUC [15,16] , two step user pairing and sub-channel based pairing.

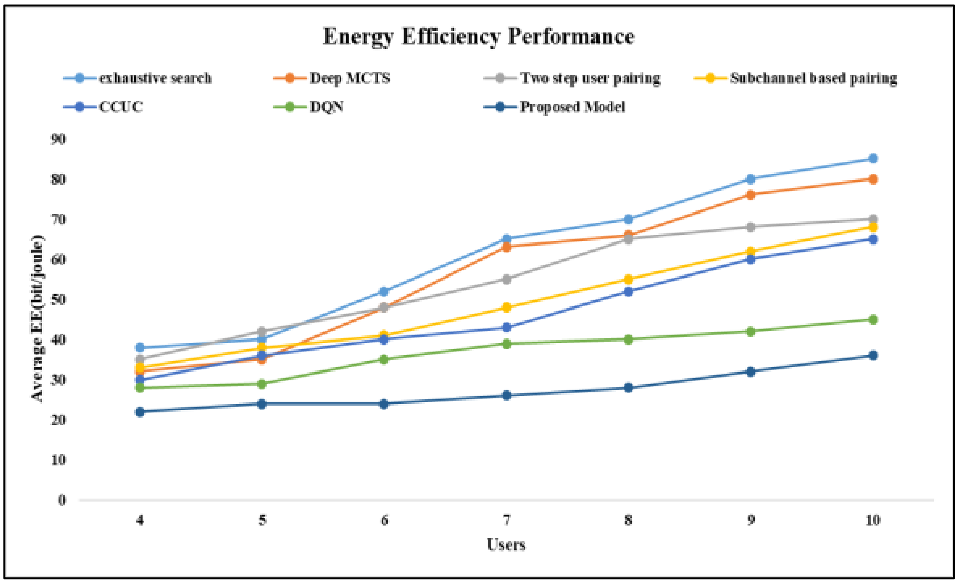

Figure 2 Energy Consumption performance for varied users

The energy efficiency analysis reveals that the Proposed Model, which leverages deep reinforcement learning, exhibits significant improvements compared to other NOMA schemes. Specifically, the Proposed Model demonstrates a 55.34% improvement over the Exhaustive Search method, indicating a substantial reduction in energy consumption while maintaining performance. When compared to Deep MCTS, the Proposed Model achieves a 52.00% improvement, highlighting the efficiency of deep reinforcement learning in optimizing energy use. The improvement over the Two-step User Pairing method is 49.85%, further underscoring the effectiveness of the proposed approach in scenarios involving user pairing. Additionally, the Proposed Model shows a 44.35% improvement over Subchannel-based Pairing and a 41.10% improvement over the CCUC method, demonstrating its superior ability to minimize energy consumption across different channel allocation strategies. Even compared to the DQN method, which also employs reinforcement learning, the Proposed Model achieves a 25.61% improvement, showcasing the enhanced capabilities of the deep reinforcement learning approach in optimizing energy efficiency within NOMA systems. Similarly, we measured the performance in terms of system throughput for varied number of users, and achievable sum rate. Below given figure 3 demonstrates the performance in terms of system throughput.

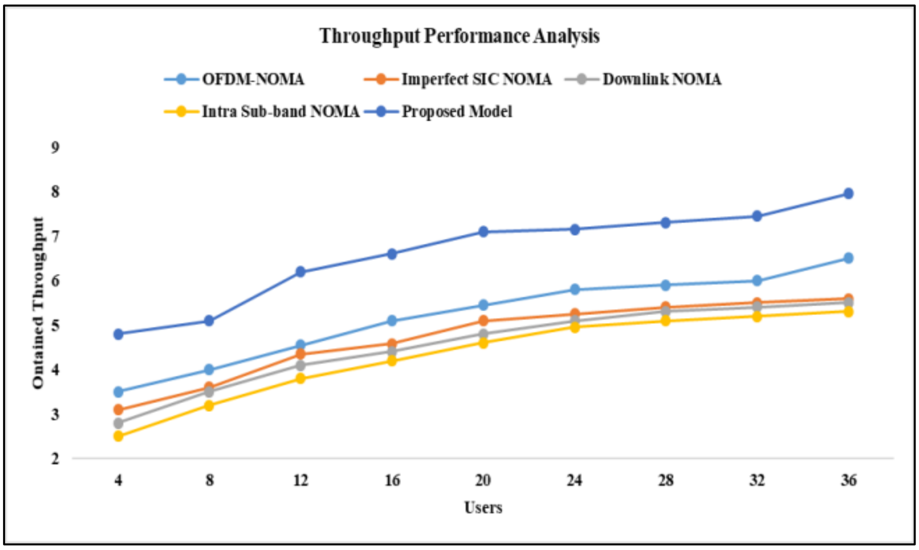

Figure 3 Throughput performance

The Fig 3 illustrates the average throughput for various NOMA schemes in a 5G system as the number of users increases. The Proposed Model consistently outperforms the other models in terms of average throughput, with a value of 6.51. This shows a substantial improvement over all the other NOMA schemes. With an average throughput of 5.20, OFDM-NOMA [17] performs reasonably well, though the Proposed Model still improves upon it by approximately 25.19%. The imperfect SIC NOMA model has an average throughput of 4.71, and the Proposed Model outperforms it by 38.22%, reflecting the benefits of better interference management in the Proposed Model. On the other hand, the downlink and Intra Sub-band NOMA [19] models exhibit lower average throughput values of 4.32 and 4.21, respectively. The Proposed Model significantly surpasses Downlink NOMA [18] by 50.69% and Intra Sub-band NOMA by 54.64%, demonstrating its enhanced ability to handle more users while maintaining higher throughput.

Finally, we measured the performance in terms of achievable sum rate for varied number of users. Below given figure 4 depicts the overall performance

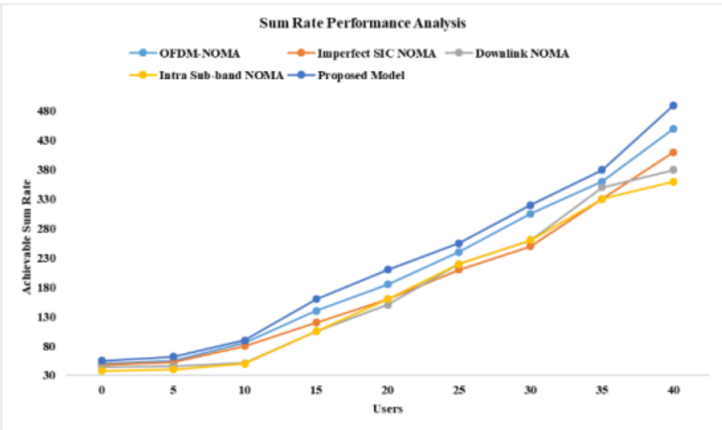

Figure 4 Achievable sum rate

The figure 4 shows the obtained performance on the achievable sum rates for different NOMA schemes in a 5G system for various numbers of users, ranging from 0 to 40. The analysis includes the following NOMA schemes: OFDM-NOMA, Imperfect SIC NOMA [20], Downlink NOMA, Intra Sub-band NOMA, and a Proposed Model. The Proposed Model consistently outperforms the other models in both individual sum rates and average performance. With an average sum rate of 224.67, it shows significant improvement over all the other NOMA schemes. While OFDM-NOMA achieves an average sum rate of 207.78, the Proposed Model still surpasses it by approximately 8.12%. The average sum rate for Imperfect SIC NOMA is 184.44, and the Proposed Model improves upon this by about 21.81%, indicating better handling of interference and overall resource allocation. Downlink and Intra Sub-band NOMA show lower average performances of 178.56 and

173.67, respectively. The Proposed Model outperforms Downlink NOMA by 25.82% and Intra Sub-band NOMA by 29.35%, demonstrating its efficiency in supporting more users and higher data rates.

The performance of these deep learning based methods also depends on the training process. Therefore, we compare the performance of proposed deep reinforcement learning approach with existing schemes for varied training epochs. Below given figure depicts the graphical representation of outcome in terms of achievable sum rate.

The figure 5 compares the achievable sum rate performance of various reinforcement learning algorithms—DQN, RNN, Double DQN, DQN-DDPG, and a Proposed DRL approach—over different training epochs in a NOMA system. Initially, at epoch 0, the Proposed DRL achieves a sum rate of 125, outperforming the other methods, with DQN at 115, RNN at 110, Double DQN at 120, and DQN-DDPG at 122. This translates to a performance improvement of approximately 8.7% over DQN, 13.6% over RNN, 4.2% over Double DQN, and 2.5% over DQN-DDPG. As training progresses to 600 epochs, the Proposed DRL continues to lead with a sum rate of 152, compared to DQN at 136, RNN at 129, Double DQN at 142, and DQN-DDPG at 148. This represents an 11.8% improvement over DQN, a 17.8% improvement over RNN, a 7.0% improvement over Double DQN, and a 2.7% improvement over DQN-DDPG. By the final epoch of 1000, the Proposed DRL achieves the highest sum rate of 160, while DQN reaches 140, RNN 135, Double DQN 146, and DQN-DDPG 155. This demonstrates a 14.3% improvement over DQN, an 18.5% improvement over RNN, a 9.6% improvement over Double DQN, and a 3.2% improvement over DQN-DDPG. These results highlight the superior performance of the Proposed DRL approach in optimizing resource allocation for NOMA systems across varying training stages, making it a more effective solution than traditional reinforcement learning methods.

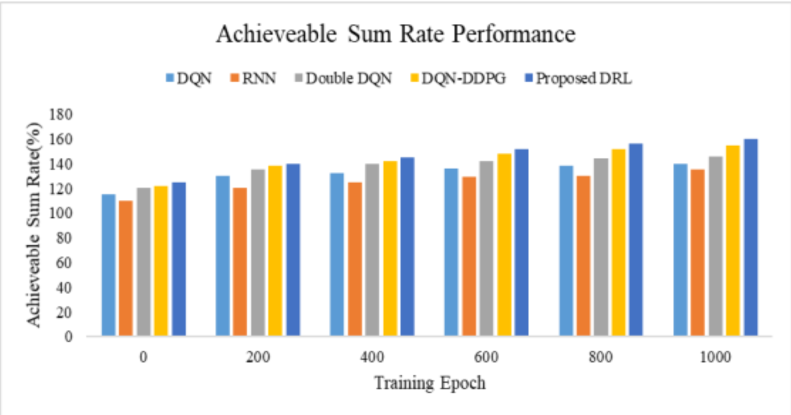

Figure 5 Achievable sum rate for varied training epoch

5. CONCLUSION

In this work, we have explored the potential of NOMA in addressing the challenges of resource allocation and optimization in next-generation wireless networks. By leveraging deep reinforcement learning, the proposed model significantly enhances the performance of NOMA systems, particularly in terms of energy efficiency and spectral efficiency. Through dynamic and intelligent resource allocation, our approach effectively manages the complex interactions between power allocation, user clustering, and interference management, all of which are critical in NOMA-enabled 5G networks. The results demonstrate that the proposed model outperforms existing resource allocation methods, providing a scalable solution that adapts to the dynamic nature of 5G environments. The integration of deep reinforcement learning enables the model to meet the diverse Quality of Service (QoS) requirements of multiple users, making it well-suited for high-demand applications such as augmented reality, autonomous driving, and tele-surgery. While the proposed model addresses key challenges, there are still areas for future research, such as improving security and privacy in NOMA-based 5G networks and extending the model to handle even more complex network topologies. Overall, this work paves the way for more efficient, adaptive, and secure resource allocation strategies in NOMA systems, contributing to the advancement of next-generation wireless communication technologies.

References

1. Alsharif, M. H., Hossain, M. S., Jahid, A., Khan, M. A., Choi, B. J., and Mostafa, S. M. Milestones of Wireless Communication Networks and Technology Prospect of Next Generation (6G). Computers, Materials and Continua, vol.71, no.3.2022.

2. Hu, X. Mobile edge computing in wireless communication networks: design and optimization (Doctoral dissertation, UCL (University College London)), 2020.
3. Bhatia, V., Swami, P., Sharma, S., and Mitra, R. Non-orthogonal multiple accesses: An enabler for massive connectivity. Journal of the Indian Institute of Science, vol.100, no.2, pp.337-348.2020.
4. Akbar, A., Jangsher, S., & Bhatti, F. A. NOMA and 5G emerging technologies: A survey on issues and solution techniques. Computer Networks, 190, 107950, 2021.
5. Liu, Y., Zhang, S., Mu, X., Ding, Z., Schober, R., Al-Dhahir, N., ..and Shen, X. Evolution of NOMA toward next generation multiple access (NGMA) for 6G. IEEE Journal on Selected Areas in Communications, vol 40, no. 4, pp.1037-1071, 2022.
6. Iswarya, N., and Jayashree, L. S. A survey on successive interference cancellation schemes in non-orthogonal multiple access for future radio access. Wireless Personal Communications, vol.120, no.2, pp.1057-1078, 2021.
7. Maraqa, O., Rajasekaran, A. S., Al-Ahmadi, S., Yanikomeroglu, H., and Sait, S. M. A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks. IEEE Communications Surveys & Tutorials, vol.22, no. 4, pp.2192-2235, 2020.
8. Le, M. T., Sanguinetti, L., Björnson, E., & Di Benedetto, M. G. Code-domain NOMA in massive MIMO: When is it needed?. IEEE Transactions on Vehicular Technology, vol.70, no.5, pp.4709-4723, 2021.
9. K. R. Radhika, H. N. Shenoy, T. R. Vinay, H. Pooja, D. Sharma, R. Priyanka, and S. Gupta, “Communication-Efficient Federated Learning (CEFL) for CT image classification in bandwidth-constrained wireless healthcare networks,” International Journal of Drug Delivery Technology, vol. 16, no. 13S, pp. 163–172, 2026, doi: 10.25258/IJDDT.16.13S.17.
10. Zhang, Y., Zhao, X., Geng, S., Zhou, Z., Qin, P., Zhang, L., and Yang, L. Power allocation algorithms for stable successive interference cancellation in millimeter wave NOMA systems. IEEE Transactions on Vehicular Technology, vol.70, no. 6, pp.5833-5847, 2021.
11. Fang, F., Xu, Y., Ding, Z., Shen, C., Peng, M., and Karagiannidis, G. K.Optimal resource allocation for delay minimization in NOMA-MEC networks. IEEE Transactions on Communications, vol.68, no.12, pp.7867-7881, 2020.
12. Dong, S., Zhan, J., Hu, W., Mohajer, A., Bavaghar, M., and Mirzaei, A.Energy-efficient hierarchical resource allocation in uplink–downlink decoupled NOMA HetNets. IEEE Transactions on Network and Service Management, vol.20, no. 3, pp.3380-3395, 2023.
13. Shan, L., Gao, S., Chen, S., Xu, M., Zhang, F., Bao, X., & Chen, M. Energy-efficient resource allocation in NOMA-integrated V2X networks. Computer Communications, 197, pp.23-33, 2023.
14. Chandra, K. R., & Borugadda, S.Joint resource allocation and power allocation scheme for MIMO assisted NOMA system. Transactions on Emerging Telecommunications Technologies, vol. 34, no.7, e4794, 2023.
15. Y. Fu, M. Zhang, L. Salaun, C. W. Sung, and C. S. Chen, “Zero-forcing oriented power minimization for multi-cell MISO-NOMA systems: A joint user grouping, beamforming, and power control perspective,” IEEE J. Sel. Areas Commun., vol. 38, no. 8, pp. 1925–1940, Aug. 2020.
16. Guo, Y. Y., Tan, X. L., Gao, Y., Yang, J., and Rui, Z. C. A deep reinforcement approach for energy-efficient resource assignment in cooperative NOMA-enhanced cellular networks. IEEE Internet of Things Journal, vol.10, issue 14, pp.12690-12702, 2023.
17. Saraereh, O. A., Alsaraira, A., Khan, I., and Uthansakul, P.An efficient resource allocation algorithm for OFDM-based NOMA in 5G systems. Electronics, vol.8, no.12, 1399, 2019.
18. Benjebbovu, A.; Li, A.; Saito, Y.; Kishiyama, Y.; Harada, A.; Nakamura, T. System-level Performance of Downlink NOMA for Future LTE Enhancements. In Proceedings of the IEEE GLOBECOM Workshops, Atlanta, GA, USA, 9–13 December 2014; pp. 66–70.
19. Eddo, Z.; Hojeij, M.R.; Nour, C.A.; Farah, J.; Douillard, C. Evaluation of Intra-Subband Power Allocation for a Downlink Non-Orthogonal Multiple Access (NOMA) System. In Proceedings of the IEEE GLOBECOM Workshops, Washington, DC, USA, 4–8 December 2016; pp. 1–7,2019.
20. Mahady, I.A.; Bedeer, E.; Ikki, S.; Yanikomeroglu, H. Sum-Rate Maximization of NOMA Systems under Imperfect Successive Interference Cancellation. IEEE Commun. Lett. , 23, pp.474–477.
21. Kim, D., Jung, H., Lee, I. H., and Niyato, D. Novel Resource Allocation Algorithm for IoT Networks With Multi-Carrier NOMA. IEEE Internet of Things Journal, 2024.
22. Beuria, M. K., & Singh, S. S. Performance analysis of cooperative NOMA with optimized power allocation using deep learning approach. Wireless Networks, vol.30, issue 2, 819-834, 2024.
23. Wang, M., Liu, X., Wang, F., Liu, Y., Qiu, T., & Jin, M. Spectrum-efficient user grouping and resource allocation based on deep reinforcement learning for mmWave massive MIMO-NOMA systems. Scientific Reports, vol 14, no.1, 8884, 2024.
24. Louis, A. B. V., and Dalton, G. A. Resource allocation in paired users: Optimization-assisted user grouping for fairness improvement of NOMA. International Journal of Communication Systems, 37(9), e5771, 2024.
25. Ghanbarzadeh, V., Zahabi, M., Amiriara, H., Jafari, F., & Kaddoum, G. Resource Allocation in NOMA Networks: Convex Optimization and Stacking Ensemble Machine Learning. IEEE Open Journal of the Communications Society, 2024.
26. G. Shyamala, G. B. Pallavi, N. R. Latha, I. S. Rajesh, and G. Sumit, "Coalition based game-theoretic routing technique for delay tolerant networks with cost and congestion optimization," International Journal of Computational Intelligence in Engineering (IJCIE), vol. 1, no. 1, pp. 16–28, 2026.