

Transformer-Based Multi-Agent Reinforcement Learning with Residual Attention Networks for Priority-Aware and Energy Optimization in UAV-assisted NOMA IoT

¹Naama Methab Athab and ²Abdullahi Abdu Ibrahim

¹Computer Engineering Department
Altinbas University, Turkey
213720618@ogr.altinbas.edu.tr

²Software Engineering Department
Altinbas University, Turkey
abdullahi.ibrahim@altinbas.edu.tr

Abstract: The huge demand for Internet of Things (IoT) devices has created a very big need for better resource sharing and energy use in the Unmanned Aerial Vehicle (UAV) networks. Normal reinforcement learning approaches fail in complex environmental conditions, and it is not able to fully capture the complex network's interactions. Also, the current priority-aware access and energy optimization systems struggle with mixed traffic patterns, sudden changes in user needs, and fairness among the devices. These limitations show the urgent requirement for stronger learning models that can deal with the evolving conditions more effectively. In this context, Non-Orthogonal Multiple Access (NOMA) plays an important role because it allows multiple IoT devices to share the same spectrum resources by separating them in the power domain. While NOMA improves spectral efficiency, it also introduces new challenges in managing interference and balancing energy consumption. Therefore, in this work, a new approach is presented that builds on the existing deep reinforcement learning framework for the UAV-assisted NOMA IoT networks. This method uses Transformer-based Multi-Agent Reinforcement Learning with Residual Attention Network (TMARL-RAN). In the network, several agents in DRL are assigned to several tasks to perform accurate decision-making. Also, the transformer is used as a feature extractor, and the RAN module improves feature extraction by utilizing residual connections with attention. This synergy allows the proposed TMARL-RAN to learn better policies, adapt faster to complex IoT scenarios, and finally provide effective decision-making strategies for energy optimization in UAV-driven NOMA IoT networks. At last, the developed model's outcomes are compared with the conventional techniques, where it provides a strong solution for the UAV-assisted NOMA IoT networks.

Keywords: UAV-assisted NOMA IoT; Priority-Aware and Energy Optimization; Transformer; Multi-Agent Reinforcement Learning; Residual Attention Networks

Received: May 15, 2026 Revised: May 28, 2026 Accepted: June 3, 2026

Nomenclature

Abbreviation	Description
IoT	Internet of Things
UAV	Unmanned Aerial Vehicle
NOMA	Non-Orthogonal Multiple Access
TMARL-RAN	Transformer-based Multi-Agent Reinforcement Learning with Residual Attention Network
MARL	Multi-Agent Reinforcement Learning
DRL	Deep Reinforcement Learning
RAN	Residual Attention Network
UE	User Equipment
QoS	Quality of Service
SIC	Successive Interference Cancellation
OMA	Orthogonal Multiple Access
RIS	Reconfigurable Intelligent Surface
IRM-TR	Iterative Resource Management-Trajectory Refinement

1. Introduction

Due to the rapid growth of IoT applications, the number of ground User Equipment (UE) has grown massively, and this puts very heavy demands on wireless networks [6]. These devices need not only continuous connectivity but also different levels of Quality of Service (QoS). In many cases, the normal terrestrial communication infrastructure is either missing or damaged, and traditional cellular systems cannot provide reliable services. To overcome this, UAVs have become a very practical solution. They can be deployed quickly and converge exactly when and where it is needed [7]. Because of these features, UAVs are now seen as cost-effective aerial

base stations that can deliver seamless wireless access to the IoT devices on the ground. At the same time, NOMA has become an important technology; it is designed to allow multiple users to share the same frequency resources at the same time. Instead of separating users by time or frequency like traditional methods [8], NOMA separates them in a power domain. This is achieved through Successive Interference Cancellation (SIC). The advantages are that spectral efficiency is improved, and massive connectivity is also supported. So, NOMA fits naturally with the UAV-assisted IoT networks, since it can handle the huge access demands of many devices. Its principles in wireless systems have also been utilized in the UAV-assisted NOMA IoT [9].

UAVs are powered by limited onboard batteries, and their flight time, coverage ability, and communication performance all depend on how efficiently this energy is used. If energy is wasted, UAVs cannot stay in the air long enough to serve IoT devices, which directly affects connectivity and reliability. So, energy optimization in UAV-assisted NOMA systems plays a crucial role. In IoT scenarios, thousands of devices may need access at the same time, and UAVs must balance between priority-aware scheduling and energy-efficient transmission [10]. Hence, researchers have already done a lot of work on optimizing UAV-assisted NOMA systems, focusing mainly on improving throughput, sum rate, and energy efficiency. The traditional model worked on a throughput maximization problem by jointly optimizing user scheduling, per-timeslot power allocation, and UAV trajectory [11]. Also, an alternating iterative algorithm was proposed to solve the non-convex subproblems, and its results showed that NOMA performs better than Orthogonal Multiple Access (OMA). Additionally, a hybrid Reconfigurable Intelligent Surface (RIS)-UAV cooperative NOMA scheme, where UAV 3D placement and power allocation were optimized to maximize system sum rate [12]. In emergency communication scenarios, a multi-UAV full-duplex NOMA network was studied, and an Iterative Resource Management-Trajectory Refinement (IRM-TR) algorithm for joint subcarrier scheduling and power allocation was proposed [13]. Moreover, a dual-layer frequency-power slicing framework was designed along with a QoS-aware convex optimization model to coordinate resources for heterogeneous services. However, most of these methods are deterministic and rely on convex optimization or alternating iteration [14]. It usually assumes that perfect static channel state information is available, which is not realistic in practice. Because UAVs are mobile and IoT devices connect randomly, the network environment is highly dynamic. These traditional methods cannot adapt well to such changes. Also, they face very high computational complexity and poor performance when dealing with large-scale problems that involve multiple UAVs, many users, and several constraints at the same time. As a result, they fail to meet the fast decision-making needs of dynamic UAV networks [15]. Therefore, there is an urgent demand for an effective energy optimization model in UAV-assisted NOMA IoT.

The major contribution of this paper can be summarized as follows:

- To design a deep learning model to improve the UAV-assisted IoT networks that manage both energy and device access. In general, it helps UAVs stay longer in the air by using energy more wisely, while at the same time making sure that devices with urgent data get faster services. Thus, the developed model can handle many IoT devices at once without wasting resources.
- To introduce TMARL-RAN, which combines advanced strategies to jointly manage energy resources and priority scheduling. Here, the UAV-enabled NOMA IoT system state is given as input to the MARL framework. Before that, a transformer module is incorporated to extract a wide range of input state features, and then RAN is used to improve feature quality. By doing so, it balances energy efficiency with priority-aware access, ensuring that critical IoT devices are served first.

Organizations: The rest of the paper is organized as follows: Section 2 reviews the existing studies on UAV-NOMA optimization, describing its key challenges and benefits. Section 3 presents the system models of NOMA with IoT and UAV channels in NOMA with IoT, including developed model descriptions. Section 4 details the TMADRK-RAN work function for energy- and priority-aware optimization. Section 5 provides simulation results and comparative analysis with existing methods. Finally, Section 6 summarizes concise insights of the proposed framework in dynamic IoT environments.

2. Existing Works

2.1 Related Works

In 2026, Wang *et al.* [1] proposed a non-convex optimization model with the goal of reducing the weighted access failure rate for devices with different priorities in multi-UAV-assisted NOMA IoT networks. Along with this, the Proximal Policy Optimization (PPO) algorithm was used for multi-head policy design.

In 2025, Xie *et al.* [2] developed a NOMA-based multi-UAV semantic cellular framework. In this setup, UAVs transmitted semantic information using NOMA, which improved spectrum utilization. An enhanced K-means algorithm was implemented to periodically match UAVs with users; along with this, a deep reinforcement learning method called Shared Dueling Double Deep Q-Network (SD3QN) was used to jointly optimize symbols, trajectories, and power allocation.

In 2026, Bulusu *et al.* [3] focused on UAV trajectory planning; for that, a trajectory optimization called the Markov Decision Process (MDP) was modelled, and a Deep Q-Network (DQN) was applied to adaptively choose UAV movements. At last, a new reward function was designed to guide UAVs toward sensor locations.

In 2026, Tong *et al.* [4] proposed an Iterative Refinement Method With Threshold-Based Rounding (IRM-TR). This algorithm refined assignments and power allocation using convex approximations and a dynamic weight tensor, then applied rounding to ensure binary values.

In 2026, Mayarakaca *et al.* [5] integrated UAV-assisted RF Energy Harvesting (RF-EH), NOMA, and Device-to-Device (D2D) communication. In this system, UAVs transmitted RF signals. Then the network is operated under a D2D framework with NOMA, and an EH time-switching protocol with dynamic power allocation is applied. For node selection, a K-Nearest Neighbor (kNN) algorithm was employed on the UAV.

2.2 Research Gaps and Challenges

In UAV-assisted NOMA IoT systems, unmanned aerial vehicles are used as flying base stations to support massive device connectivity. The main idea is to improve the coverage, reduce latency, and balance energy usage while handling different IoT devices with varying priority levels. Traditional models try to balance energy efficiency and priority awareness but often struggle when the environment becomes complex and dynamic. Some of the issues, which are faced by the conventional model are given below:

- Traditional models often assign transmission power and bandwidth in a fixed way. This rigid allocation cannot adapt well when IoT devices suddenly change their priority.
- As the number of IoT devices grows, the optimization becomes too complex for static models. They fail to handle large-scale networks efficiently.
- UAVs have restricted battery life, but traditional models do not fully optimize energy usage during flight and communication. This leads to faster battery drain and reduced mission time.
- IoT networks are highly dynamic, with devices joining or leaving, and UAVs moving in uncertain airspace. Traditional models cannot quickly adapt to these changes.
- Devices with urgent or critical data may not get proper priority because the models lack fine-grained scheduling mechanisms.

Thus, the traditional energy optimization model faces the above challenges. Furthermore, its inability to adapt to changing environments makes it less effective. Therefore, while it sets the foundation, more flexible and advanced approaches are required to achieve true efficiency in UAV-IoT systems

3. Systematic View and its Channel Model for Developing Priority-aware Optimization

3.1 System Model of NOMA with IoT

Consider an IoT wireless network where a Base Station (BS) serves k IoT through sub-channels [16] [17]. The scenario is outdoors, with omnidirectional antennas at both the BS and IoT devices. Each wireless link follows Rayleigh fading. Importantly, here each sub-channel can support multiple IoT devices at the same time using NOMA.

Signal Model: If t_m is the set of IoT devices on sub-channel $m \in M$, and $x_{l,m}$ is the signal device $l \in S_m$, then the received signal at device $k \in S_m$ is provided in Eq. (1).

$$d_{k,m} = h_{k,m} \sum_{l=1}^{s_m} \sqrt{P_{l,m}} x_{l,m} + v_{k,m} \quad (1)$$

The channel coefficient at device k on sub-channel m , transmit power of device l is $P_{l,m}$, data signal of device l is $x_{l,m}$, and an additive white Gaussian noise with zero mean and variance ρ^2 is $v_{k,m}$.

SIC Decoding: Since multiple IoT devices share one sub-channel, SIC is applied.

→ Device k can decode signals of device l if $|h_{l,m}|^2 \leq |h_{k,m}|^2$.

→ To reduce decoding complexity, each sub-channel can host at most $s_{\max}$ devices that is. ..

Achievable Rate: The achievable rate of device k on a sub-channel m is presented in Eq. (2).

$$R_{k,m} = \log_2(1 + \xi_{k,m}) \quad (2)$$

The Signal-to-Interference-Noise Ratio (SINR) is $\xi_{k,m}$ that is described in Eq. (3).

$$\xi_{k,m} = \frac{p_{k,m} |h_{k,m}|^2}{|h_{k,m}|^2 \sum_{l=1}^{K-1} P_{l,m} + \rho^2} \quad (3)$$

SIC Constraint: For SIC to be successful, the minimum power gap between devices must satisfy Eq. (4).

$$P_{k,m} |h_{k-1,m}|^2 - \sum_{l=1}^{k-1} P_{l,m} |h_{k-1,m}|^2 \geq \beta \quad (4)$$

The minimum required power gap for reliable SIC is β . Thus, this is how a BS serves multiple IoT devices using NOMA. The pictorial representation of this system model is shown in Fig. 1.

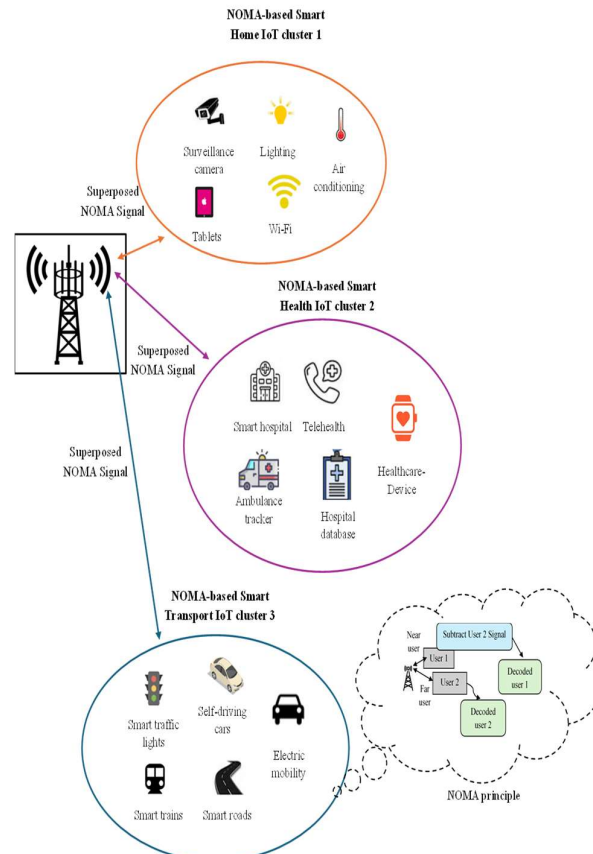


Fig. 1. System model of NOMA-enabled IoT network

3.2 UAV Channel in NOMA-IoT

In a NOMA-IoT system, UAVs are used as an aerial BS to provide uplink access for the ground IoT devices [1]. Unlike the fixed BS, the UAV can move freely in 3D space in dense areas. Therefore, by combining UAV channels with NOMA, the system achieves better and more efficient spectrum use for the IoT applications.

Channel Model Setup:

Ground Devices: Let the set of IoT devices be $N = \{1, 2, \dots, n\}$. Each device n has a 2D position $\theta_n = (x_n, y_n)$.

UAVs: A cluster of UAVs is denoted by $T = \{1, 2, \dots, t\}$. The 3D position of the UAV t at time slot u is defined in Eq. (5).

$$\theta_t(s) = (x_t(s), y_t(s), h_t(s)) \quad (5)$$

The altitude is denoted as $h_t(s)$.

UAV Mobility: UAV movement is described using spherical coordinates such as Azimuth and Elevation angle. At each time slot, the UAV samples φ^* and ϕ^* to update its position. This process is illustrated in Eq. (6).

$$\begin{aligned} x_t(s+1) &= x_t(s) + v_t(s) \cos(\phi^*) \cos(\varphi^*) \\ y_t(s+1) &= y_t(s) + v_t(s) \cos(\phi^*) \sin(\varphi^*) \\ h_t(s+1) &= h_t(t) + v_n(t) \sin(\varphi^*) \end{aligned} \quad (6)$$

Here, the UAV step length is $v_t(s)$.

Propulsion power: For rotary-wing UAVs, propulsion power consumption is modelled in Eq. (7).

$$P(V) = P_0 \left(1 + \frac{3v^2}{U_{tip}^2} \right) + P_i \left(\sqrt{1 + \frac{v^4}{4v_0^2}} - \frac{v^2}{2v_0^2} \right) + \frac{1}{2} d_0 \rho s A v^3 \quad (7)$$

The blade profile power is P_0 , induced power during hovering is P_i , rotor tip speed is U_{tip}^2 , induced velocity in hover is v_0 , and fuselage drag ratio, rotor reliability, air density, and rotor disk area are denoted as $d_0 \rho s A$.

Received Signal Model: The received signal at UAV is denoted in Eq. (8).

$$y_t(s) = \sum_{m \in M_{n(t)}} h_{m,n}(s) \sqrt{P_m} x_m + n_s \quad (8)$$

The set of IoT devices currently communicating with UAV is $M_{n(t)}$, composite channel gain is $h_{m,n}(s)$, and n_s is the noise term. Hence, by combining UAV mobility with NOMA's power domain multiplexing, the system achieves higher connectivity for IoT networks.

3.3 Proposed Model and its Explanation

In traditional UAV-assisted NOMA IoT networks, optimization has always been a very big challenge. Most of the earlier models often relied on static methods or heuristic algorithms that cannot adapt well to a fast-changing environment. For example, UAVs have limited battery power; their communication channels can vary with movement, and IoT devices generate requests with different priority levels. When these factors change quickly, traditional models struggle to balance energy usage with high throughput. It often ends up wasting energy by flying inefficient trajectories and failing to serve critical IoT devices on time. This leads to poor scheduling, unstable resource allocation, and reduced quality of service.

The proposed TMARL-RAN model is designed to overcome these limitations by combining three powerful components: transformer-based state encoding, residual RAN for feature processing, and MADRL. The workflow begins with the UAV-IoT system providing input attributes. These inputs form the basic state representation. First, the transformer encoder is applied to handle the sequential nature of the state data. Since UAV conditions and IoT requests change over time, the transformer injects positional information using sine and cosine functions. This allows the model to understand the order of events and capture temporal dependencies. Then, Multi-Head Self-Attention (MHSA) is used to focus on the most relevant historical states. This ensures that the model does not just react to the present but also learns from the past, which is crucial for UAV decision-making. Next, the RAN refines the features extracted by the transformer. The trunk branch processes the raw features, while the mask branch generates

attention masks that highlight important signals. Through residual attention learning, the model combines these two outputs in a way that preserves original information while still emphasizing critical features. This prevents feature loss and ensures that UAVs focus on urgent IoT requests and energy-related signals. After feature extraction and refinement, the processed state is passed into the MADRL framework. Here, multiple agents cooperate to make decisions about UAV trajectory, transmit power, CPU scaling, and priority scheduling. Each agent observes the environment, selects actions, and receives rewards based on throughput and energy cost. The actor network generates deterministic actions, while the critic network evaluates them. Target networks and replay buffers are used to stabilize training, ensuring that agents learn consistent policies even in non-stationary environments. Finally, the output of TMARL-RAN is an optimized policy that balances energy efficiency with priority-aware access. UAVs can operate longer, allocate resources fairly, and serve critical IoT devices first. Therefore, compared to the traditional models, TMARL-RAN adapts faster, learns better strategies, and provides reliable performance in dynamic UAV-IoT scenarios. The structural flow of the proposed model is explained in Fig. 2.

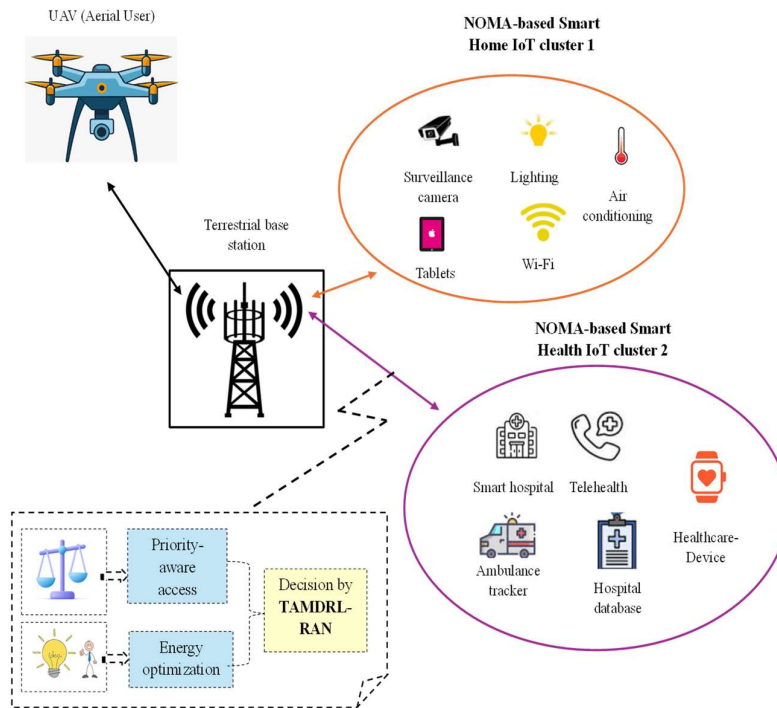


Fig. 2. Architecture of the developed model for energy- and priority-aware access optimization in UAV-enabled NOMA IoT systems

4. Proposed Transformer with Multiagent RL for Priority-aware Optimization

4.1 Residual Attention Network

RAN [18] is a deep learning architecture that combines the strength of residual learning with attention mechanisms. The main idea is to guide feature extraction by focusing on the most important parts of the input while suppressing irrelevant or noisy information.

Structure of RAN: It has two main branches, such as the trunk branch (for feature processing) and the mask branch (for generating attention maps).

At first, the trunk branch processes the input features using residual units. Then the mask branch learns a soft attention mask $N(x)$. It highlights the useful features and suppresses noise. The final output of an attention module is given in Eq. (9).

$$H_{i,c}(x) = N_{i,c}(x), T_{i,c}(x) \quad (9)$$

The output feature at spatial position i and channel c is $H_{i,c}(x)$, the attention mask value is $N_{i,c}(x)$ that ranges between 0 and 1, and the trunk branch feature is $T_{i,c}(x)$. This shows how the mask acts like a filter, weighting the trunk features before passing them forward.

Attention residual learning: Simply stacking attention modules can weaken features due to the repeated multiplication with masks. To solve this, attention residual learning modifies the output, that is shown in Eq. (10).

$$H_{i,c}(x) = (1 + N_{i,c}(x)) \cdot F_{i,c}(x) \quad (10)$$

The term $(1 + N_{i,c}(x))$ ensures that even if the mask is close to zero, the original features are preserved and the original trunk feature is pointed as $F_{i,c}(x)$. Thus, RAN provides a powerful way to enhance feature extraction by combining residual learning with adaptive attention. The structure of the general RAN model is depicted in Fig. 3.

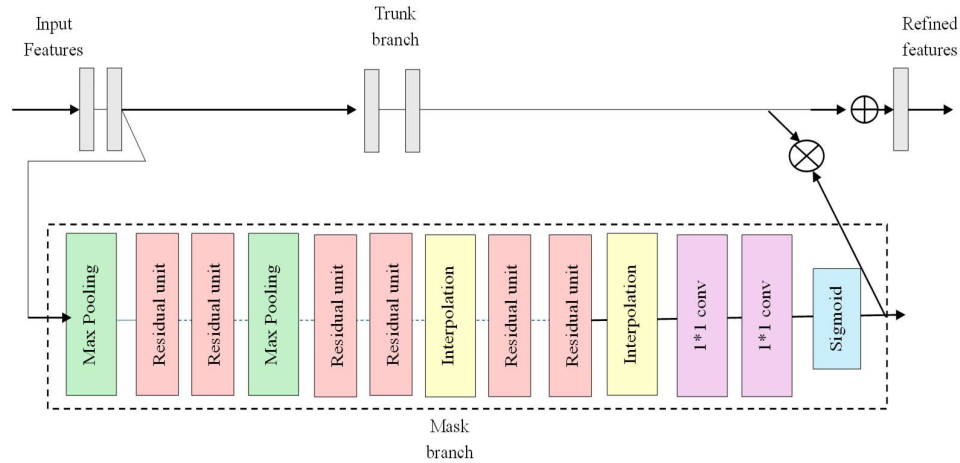


Fig. 3. RAN model

4.2 Recommended TMARL-RAN

Purpose of TMARL-RAN: The main purpose of the proposed TMARL-RAN model is to provide a reliable and efficient solution for UAV-assisted NOMA IoT networks. It goals is to balance energy optimization with priority-aware access, so UAVs can stay longer in operation while still serving critical IoT devices first. This model is designed to deliver better decision-making strategies compared to conventional optimization methods.

The workflow of the TMARL-RAN is illustrated as follows:

Input to the TMARL-RAN system: The TMARL-RAN receives certain input attributes from the UAV-enabled NOMA IoT networks, which include device requests, UAV channel conditions, UAV energy status, and priority levels of the IoT data. These inputs form the basis for the decision-making process.

MADRL-Learning Framework [19]: In the TMARL-RAN framework, the initialized input attributes are given as input to the MADRL, where it works in the belonging process (1-8 steps):
1. Environment and state representation: At each time step t , the environment is described by a state vector es_t , as explained in Eq. (11).

$$es_t = \{b_p, f_c, c_h\}_{i=1}^N \quad (11)$$

The remaining battery power of UAV (i^{th}) is b_p , CPU frequency of UAV is f_c , channel condition between UAV, IoT devices is c_h , and the number of UAVs in the system is N . Thus, this state captures the energy status, computation capability, and communication quality of the UAVs.

Transformer-based state feature extraction: Since raw state data is sequential and constantly changing, directly learning policies from it can lead to poor performance. To overcome this, TMARL-RAN employs a transformer encoder [20] before the policy learning. It uses positional encoding to process the input state data, illustrated in Eq. (12) and Eq. (13).

$$F(pos, 2b) = \sin\left(\frac{v}{10000^{2t/m}}\right) \quad (12)$$

$$F(pos, 2b+1) = \cos\left(\frac{u}{10000^{2t/m}}\right) \quad (13)$$

Here, pos is the time step, b is the dimension index, and m is the overall model dimension. The positioned input sequence is then processed by Multi-Head-Self-Attention, explained in Eq. (14).

$$A(m, n, o) = \text{softmax}\left(\frac{mn^T}{\sqrt{d_k}}\right)O \quad (14)$$

The query, key, and value matrices are m, n, o , and the key dimension is d_k . This process ensures that the transformer extracts the high range of input features. The final features obtained from the transformer module is denoted as X_t .

RAN-based feature refining: To refine the features X_t obtained from the transformer module, the RAN is employed in the developed TMARL-RAN system. Here, the RAN uses two distinct branches trunk and the mask branch. The trunk branch processes the input features X_t using residual blocks. Then the mask branch generates an attention mask, that acts like a filter for the refining process. At last, the trunk and mask outputs are combined to produce the final enriched feature representation (stated in Eq. (9)). Then, to ignore weakening features, the RAN uses residual attention learning (Eq. (10)). Finally, the obtained refined feature from the RAN is utilized by the actor and critic networks of the MADRL for further decision-making process.

2. Actor network: It generates deterministic action from the current feature state, described in Eq. (15).

$$a_t = \mu(es_t \setminus \gamma_\phi) \quad (15)$$

The actor network parameter is γ_ϕ .

3. Action space: Each agent selects an actions $a_t \in A$, which includes continuous adjustments such as transmit power P_i , flight trajectory T_i , and CPU scaling frequency F_i . Thus, the action vector is defined in Eq. (16).

$$a_t = \{P_i, T_i, F_i\}_{i=1}^N \quad (16)$$

4. Reward: The immediate feedback after taking action a_t is rewarded, shown in Eq. (17).

$$r_t = F_t - \phi_t \quad (17)$$

The amount of data successfully delivered is F_t and the operational cost is ϕ_t .

5. Replay buffer storage: It stores the past transitions of the first executed actions and helps the agents learn from experience, then prevents instability by random sampling by mini-batches.

6. Critic network: It measures the error between the predicted Q-value and the target. This operation is explained in Eq. (18).

$$C(\gamma_F) = E_\phi[F(es, a \setminus \gamma_F) - h_t]^2 \quad (18)$$

The critic's output is $F(es, a \setminus \gamma_F)$ and target signal is h_t .

7. Target network update: To stabilize training, parameters are updated slowly based on the actor network's actions. This process is summarized in Eq. (19).

$$\begin{aligned} \theta'_Q &\leftarrow \tau\theta_Q + (1-\tau)\theta'_Q \\ \gamma'_\phi &\leftarrow \tau\gamma_Q + (1-\tau)\gamma'_Q \end{aligned} \quad (19)$$

The estimated Q-value by the critic network is θ_Q and the target network is θ'_Q , deterministic actions generated by the actor network is γ_Q and target network is γ'_Q .

3. Next state transition probability: The environment evolves after each action. The probability of moving to the next state is explained in Eq. (13).

$$P(es_{t+1} | es_t, a_t) \quad (20)$$

Probability of moving to the next state is pointed as es_{t+1} .

Outcomes from the TMARL-RAN: Therefore, by using the above strategies, the TMARL-RAN presents a balanced decision strategies for energy efficiency and priority-aware access. This also ensures that UAVs can operate longer while reliably handling vital IoT devices. This makes the novel TMARL-RAN more suitable for UAV-assisted NOMA IoT. The overall operation flow of the TMARL-RAN is explained via a figure in Fig. 4.

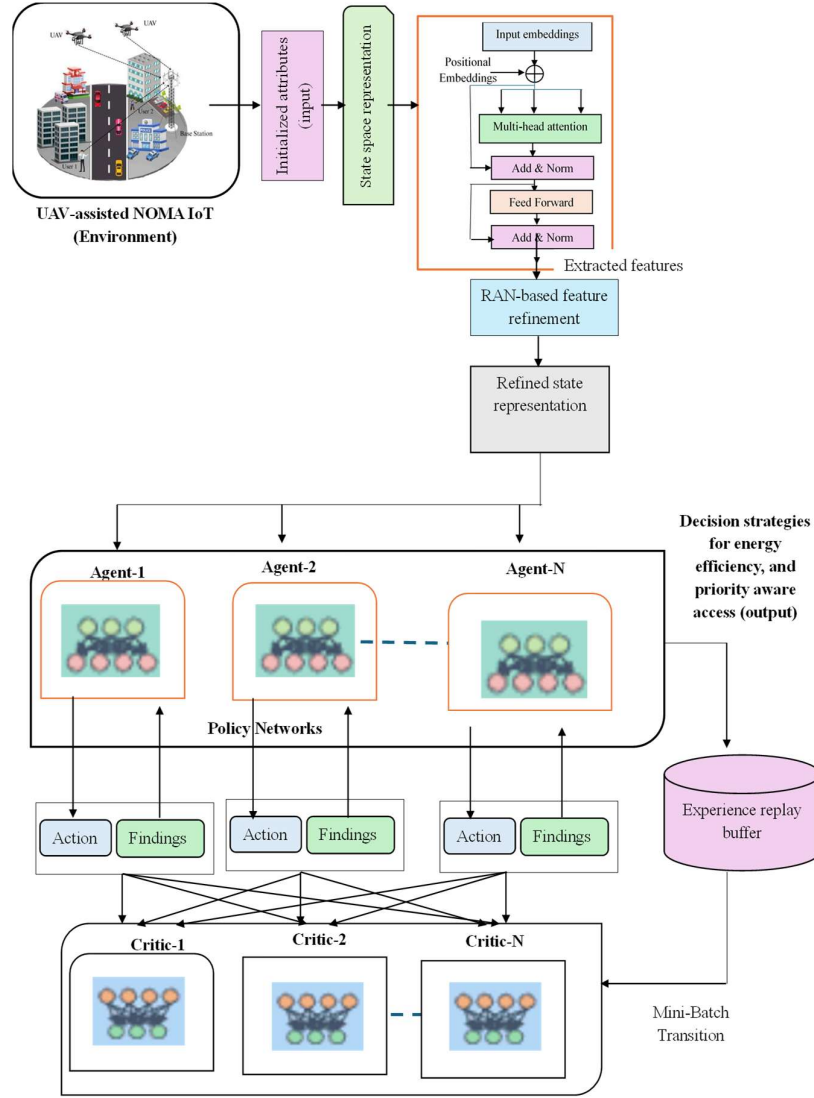


Fig. 4. Workflow of the recommended TMARL-RAN framework

5. Results and Discussions

5.1 Simulation Setup

The proposed model was implemented and executed in the Python platform, and then, to evaluate the decision-making ability of the proposed framework, certain experiments were conducted with the conventional approaches. This includes methods such as DQN [3], PPO [1], Deep Deterministic Policy Gradient (DDPG) [21], and Deep Reinforcement Learning (DRL) [22], respectively.

5.2 System efficiency and resource optimization performance of the developed TMADRL-RAN

To evaluate the performance of TMADRL-RAN in decision-making for energy optimization and priority-aware access in UAV-assisted NOMA IoT networks, a series of experiments were

conducted comparing the proposed model with standard reinforcement learning methods, as shown in Fig. 5. The results demonstrate that TMADRL-RAN achieves superior energy efficiency, outperforming DQN by 19.4%, PPO by 12.5%, DDPG by 6.9%, and DRL by 2.7%, as illustrated in Fig. 5 (a). Likewise, the resource utilization rate of TMADRL-RAN in Fig. 5 (d) shows significant improvement, surpassing DQN by 45%, PPO by 29%, DDPG by 15%, and DRL by 6.2%. In addition, the proposed TMADRL-RAN consistently delivers stable outcomes across all other metrics. These findings clearly indicate that TMADRL-RAN is more effective for energy and priority-aware access optimization in UAV-powered NOMA IoT networks.

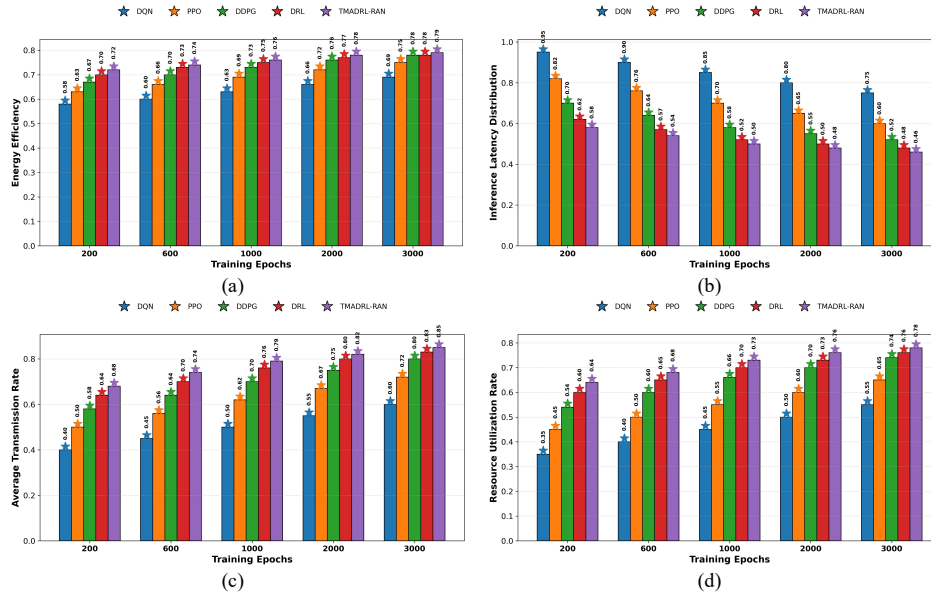


Fig. 5. Comparative performance of TMADRL-RAN versus conventional reinforcement learning methods with respect to (a) Energy Efficiency, (b) Inference Latency Distribution, (c) Average Transmission Rate, (d) Resource Utilization Rate

5.3 Network performance and policy evaluation assessment of the novel TMADRL-RAN

The efficacy of the TMADRL-RAN model was also investigated using policy evaluation measures, as shown in Fig. 6. In Fig. 6 (b), the developed TMADRL-RAN achieves a higher success rate than all other methods, outperforming DQN by 28%, PPO by 20%, DDPG by 12%, and DRL by 5.7%. This indicates that TMADRL-RAN enables more devices to successfully connect to the network without delay. Additionally, the convergence rate of TMADRL-RAN in Fig. 6 (c) demonstrates stable results, surpassing DQN by 21%, PPO by 11.5%, DDPG by 5.2%, and DRL by 2.1%. These findings highlight the faster training efficiency of TMADRL-RAN. Overall, the proposed TMADRL-RAN proves to be more suitable for UAV-assisted NOMA IoT networks.

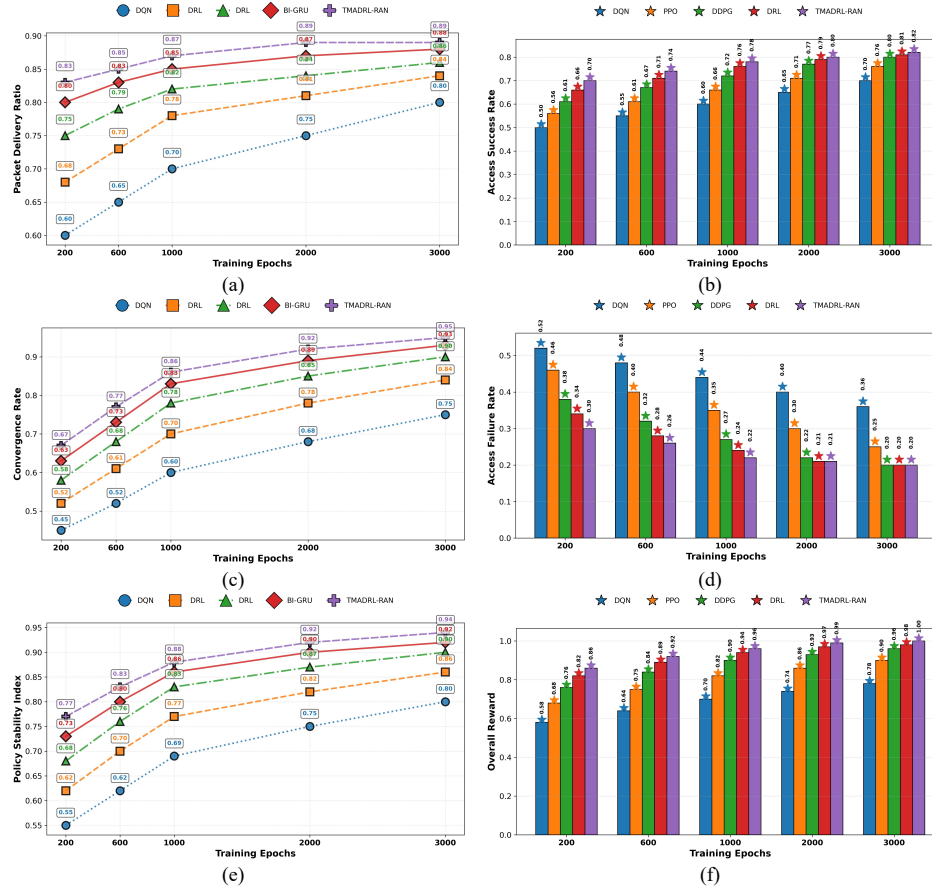


Fig. 6. Evaluation of the TMADRL-RAN compared with conventional methods for (a) Packet Delivery Ratio, (b) Access Success Rate, (c) Convergence Rate, (d) Access Failure Rate, (e) Policy Stability Index, and (f) Overall Reward

6. Conclusion

In this work, a new method called TMARL-RAN for UAV-assisted NOMA IoT networks was introduced. This approach was built on deep reinforcement learning, and it combined transformer feature extraction with the RAN to improve the learning process. By using multiple agents, the system was able to handle different tasks and make more accurate decisions. Here, the transformer helped in capturing useful features, while the RAN made the extraction process stronger. Because of this synergy, the model learned better policies and adapted faster to the complex IoT scenarios. The results showed that the implemented model gave more effective outcomes than the older techniques. For instance, the access failure rate of the TMADRL-RAN attains a lower value, such as 42% of DQN, 34% of PPO, 21% of DDPG, and 11% of DRL, respectively. Therefore, the TMARL-RAN proved that combining these advanced learning tools was effective and suitable for UAV-enabled NOMA IoT.

Future focus: Although the TMARL-RAN model showed strong performance, some points need to be focused on for future developments. First, the integration of the optimization algorithm could help speed up convergence and improve the network's training time. Then, adding attention inside the policy and critic networks of the MADRL would make the decision process more reliable. At last, incorporating explainable-based results could make the outcomes more transparent. These modifications will help TMARL-RAN to address issues such as poor convergence, the limitation in agents' ability to capture deeper features during decision making, and the lack of transparency in the decision results.

References

- [1] Wang, Z., Ding, J., Li, H., Gao, S., and Wang, Q., "Deep Reinforcement Learning Based Priority-Aware Joint Access and Energy Optimization in UAV-Assisted NOMA IoT Networks", *Physical Communication*, pp. 103179, 2026.
- [2] Xie, P., Chen, Q., Wu, J. Y., Gao, X., Xing, L., Zhang, Y., and Sun, H., "NOMA-based intelligent resource allocation and trajectory optimization for multi-UAVs assisted semantic communication networks", *Ad Hoc Networks*, vol. 171, pp. 103762, 2025.
- [3] Bulusu, C. K., Swami, P., Pattanaik, M., Mishra, M. K., and Ratnarajah, T., "NOMA-Enabled UAV Trajectory Optimization Using Deep Reinforcement Learning for Efficient IoT Data Collection", *IEEE Access*, vol. 14, pp. 48212–48221, 2026.
- [4] Wang, T., and You, C., "Joint subcarrier assignment and power allocation for UAV-assisted full-duplex NOMA emergency communication networks", *Journal of Communications and Networks*, 2026.
- [5] Mayarakaca, M. C., and Lee, B. M., "Performance analysis of UAV-assisted RF-EH and NOMA for D2D wireless sensor networks with kNN-based user selection", *Journal of Communications and Networks*, 2026.
- [6] Ren, Y., Wang, J., Xiao, Y., Zhang, X., Jiang, F., Wang, J., Jiang, J., and Lu, G., "Energy Efficiency Optimization for UAV Distribution and Resource Allocation in NOMA and Multi-UAV Assisted Wireless Networks", *IEEE Open Journal of the Communications Society*, 2025.
- [7] Ullah, R., Sarfraz, M., Sohail, M. F., Alam, S., Ghauri, S. A., and Baig, N. A., "Maximizing energy efficiency through joint resource optimization in NOMA-assisted UAV-enabled next-generation wireless networks", *Cluster Computing*, vol. 28, no. 9, pp. 581, 2025.
- [8] He, J., Shi, S., Xue, J., and Wei, S., "Resource allocation and trajectory optimization for NOMA-based and UAV-assisted satellite IoT", *Wireless Networks*, vol. 32, no. 1, pp. 59–70, 2026.
- [9] Li, M., Ku, Y., Zhu, L., Zhou, H., and Wang, H., "NOMA-Based Energy Efficiency Optimization for UAV-Assisted Relay Communication Networks", *Computer Networks*, pp. 111946, 2025.
- [10] Liu, J., Hu, H., Bai, X., Li, G., Zhang, X., Zhou, H., Li, H., and Liu, J., "Multi-UAV-assisted task offloading and trajectory optimization for edge computing via NOMA", *Sensors*, vol. 25, no. 16, pp. 4965, 2025.
- [11] Lin, X., Wang, Z., Zheng, Q., and Zhang, Z., "Energy Consumption Optimization for NOMA-Based RIS-Assisted UAV-Enabled MEC Systems", *Drones*, vol. 10, no. 6, pp. 402, 2026.
- [12] Alharbi, T. E., "Resource Allocation and Energy Harvesting in UAV-Assisted Full-Duplex Cooperative NOMA Systems", *Mathematics*, vol. 13, no. 21, pp. 3544, 2025.
- [13] Zhao, S., Gong, S., Gu, B., Li, L., Lyu, B., Hoang, D. T., and Yi, C., "Exploiting NOMA transmissions in multi-UAV-assisted wireless networks: From aerial-RIS to mode-switching UAVs", *IEEE Transactions on Wireless Communications*, vol. 24, no. 3, pp. 2530–2544, 2025.
- [14] Mohammadi, H., Shayesteh, M. G., Kalbkhani, H., and Khazali, A., "Energy Consumption Minimization in STAR-RIS-Assisted UAV-MEC Networks With NOMA", *IEEE Transactions on Green Communications and Networking*, 2026.
- [15] Sobhi-Givi, S., Nouri, M., Shayesteh, M. G., Behrooz, H., Kwon, H. H., and Piran, M. J., "Efficient optimization in RIS-assisted UAV system using deep reinforcement learning for mmWave-NOMA 6G communications", *IEEE Internet of Things Journal*, 2025.
- [16] Hamedoon, S. M., Chattha, J. N., and Bilal, M., "A novel user clustering and efficient resource allocation in non-orthogonal multiple access for IoT networks", *PLoS One*, vol. 19, no. 9, pp. e0309695, 2025.
- [17] Khan, W. U., Jameel, F., Jamshed, M. A., Pervaiz, H., Khan, S., and Liu, J., "Efficient power allocation for NOMA-enabled IoT networks in 6G era", *Physical Communication*, vol. 39, pp. 101043, 2020.
- [18] Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., and Tang, X., "Residual attention network for image classification", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3156–3164, 2017.
- [19] Andreou, A., Mavromoustakis, C. X., Markakis, E., Bourdena, A., and Mastorakis, G., "UAV-assisted IoT network framework with hybrid deep reinforcement and federated learning", *Scientific Reports*, vol. 15, no. 1, pp. 37107, 2025.
- [20] Li, F., Lang, S., Tian, Y., Hong, B., Rolf, B., Noortwyck, R., Schulz, R., and Reggeline, T., "A Transformer-Based Deep Reinforcement Learning Approach for Dynamic Parallel Machine Scheduling Problem with Family Setups", *Journal of Intelligent Manufacturing*, vol. 36, no. 7, pp. 4735–4768, 2025.
- [21] Talukdar, N., Raghav, A., Hazra, A., Barman, D. C., and Mazumdar, N., "A deep deterministic policy gradient method for optimizing task completion time and energy efficiency in UAV-assisted IoT networks", *IEEE Internet of Things Journal*, 2025.
- [22] Khairy, S., Balaprakash, P., Cai, L. X., and Cheng, Y., "Constrained deep reinforcement learning for energy sustainable multi-UAV based random access IoT networks with NOMA", *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 4, pp. 1101–1115, 2020.