

An Explainable Multi-Agent Reinforcement Learning Framework with Sparse Attention Mapping for Energy-Efficient UAV Trajectory and Resource Optimization

¹Dr. S. Satheesh Kumar and ²M. Ashok Nagasai

¹Associate Professor,
Department of CSE-AI,

Madanapalle Institute of Technology and Science (Deemed to be University),
Kadiri Raod, Angallu, Madanapalle, AP.
Pincode: 517325

drsatheeshkumars@mits.ac.in

²Assistant Professor

Dept of CSE -AI&ML

Kallam Haranathareddy Institute of Technology, Guntur.
ashok.nagasai@gmail.com

Abstract: Unmanned Aerial Vehicles (UAVs) are important in resource management for fifth-generation (5G) and wireless networks. It provides several benefits, including wider network coverage, quick deployment when needed, and flexible connectivity to handle sudden increases in data traffic. However, UAVs have limited battery power that makes energy management a major challenge. It also experiences interference from nearby UAVs that use the same frequency spectrum, which reduces the system's Energy Efficiency (EE). In addition, maintaining real-time communication with ground networks and ensuring reliable data transmission make resource management more difficult. Conventional resource management methods are not effective in optimizing energy usage, leading to higher operating costs because UAVs require frequent battery recharging. To address these issues, this paper introduces a Sparse Attention Mapping-based Explainable Multiagent Reinforcement Learning (SAM-EMRL) for optimizing EE in UAV-assisted networks. The proposed framework integrates sparse attention mapping with multi-agent reinforcement learning to enable efficient coordination among multiple UAVs while reducing computational complexity. The sparse attention mechanism selectively focuses on the most relevant network states and communication links, which allows each UAV to prioritize critical information and ignore redundant data during decision-making. The multi-agent reinforcement learning component enables UAVs to collaboratively learn optimal flight trajectories, resource allocation, and communication strategies by interacting with a dynamic network environment. In addition, the explainability module provides interpretable attention maps that highlight the factors influencing the agent's decisions, thereby improving transparency and reliability of the learning process. Therefore, the proposed SAM-EMRL framework enhances EE while maintaining stable network performance in dynamic operating conditions. Finally, validation is performed to prove the effectiveness of the proposed model.

Keywords: Unmanned Aerial Vehicles; Energy Efficiency; Sparse Attention Mapping; Explainable Multiagent Reinforcement Learning

Received: April 18, 2026 Revised: May 24, 2026 Accepted: June 25, 2026

Nomenclature

Abbreviation	Description
UAV	Unmanned Aerial Vehicle
EE	Energy Efficiency
SAM	Sparse Attention Mapping
EMRL	Explainable Multi-Agent Reinforcement Learning
SAM-EMRL	Sparse Attention Mapping-based Explainable Multi-Agent Reinforcement Learning
MARL	Multi-Agent Reinforcement Learning
ABS	Aerial Base Station
QoS	Quality of Service

1. Introduction

UAVs are widely used in various applications like search and rescue operations, emergency response, real-time surveillance, agricultural monitoring, military missions, edge computing, and wireless communication [6]. It provides cost-effective and efficient solution for improving network performance by increasing capacity, restoring connectivity during network failures caused by device malfunctions or natural disasters, and meeting the growing demand for high-speed data services [7]. UAVs also extend network coverage in densely populated, remote, and

marine environments. One of the most important applications of UAVs is their use as Aerial Base Stations (ABSs), which can be deployed quickly and autonomously to provide temporary or additional network coverage [8]. Although a single UAV can improve communication performance, its limited communication, computing, and storage capabilities make it unsuitable for mission-critical applications and networks with a large number of ground users [9]. Therefore, deploying multiple UAVs is a better approach because it provides wider coverage, higher reliability, increased network capacity, and improved quality of service in challenging communication environments [10].

In a multi-UAV communication system, multiple UAVs can work together to provide network services for ground users, which results in improved coverage and network performance. However, many existing methods depend on a centralized ground controller to coordinate UAV operations, which requires continuous information exchange between UAVs and the controller [11]. This approach is not suitable for emergencies because it increases communication overhead and makes it difficult to monitor the constantly changing locations of ground users. To ensure reliable and continuous connectivity for mobile users, UAVs need intelligent methods to optimize their flight paths while maintaining effective network coverage [12]. Since UAVs have limited battery power, it also faces challenges such as interference from nearby UAVs and other wireless access points operating on the same frequency, which reduces the EE of the system [13]. Recently, machine learning methods, especially Multi-Agent Reinforcement Learning (MARL), have been widely explored to improve EE and optimize UAV deployment. However, most of these methods assume a fixed network environment that makes them less effective in real-world scenarios where network conditions and user locations frequently change [14]. Therefore, there is a need for more adaptive and practical optimization approaches that can respond to dynamic wireless environments. This research focuses on developing an energy-efficient optimization method for UAV-assisted wireless networks to improve data processing and information transmission while adapting to changing network conditions [15].

The main contribution of the work is given below.

- ✓ This work develops an intelligent framework that manages UAV operations, communication resources, and energy consumption. The proposed approach focuses on optimizing UAV trajectories and resource allocation to reduce energy usage by maintaining reliable communication and high-quality services. The framework improves energy utilization, enhances user connectivity, and provides reliable data transmission in dynamic environments.
- ✓ To improve EE in UAV-assisted wireless networks, SAM-EMRL is proposed. The proposed method enables multiple UAVs to collaboratively learn optimal policies without relying on a central controller. The sparse attention mechanism helps UAV agents to focus on crucial network information that enhances learning efficiency and decision accuracy. The explainable reinforcement learning strategy enhances transparency and reliability of UAV decisions.

Organization: Section 2 gives the literature survey of conventional models. The system model of the designed method is given in Section 3. Section 4 describes the structural explanation of the designed method. The results and assessment of the proposed model are described in Section 5. Section 6 concludes the work.

2. Literature survey

2.1 Related Works

In 2023, Omoniwa *et al.* [1] implemented a direct collaborative communication-enabled multi-agent decentralized double deep Q-network (CMAD-DDQN) approach to improve EE in UAV-assisted wireless networks. The results demonstrated that the designed approach maximized system EE by maintaining the communication performance of the network.

In 2024, Wang *et al.* [2] proposed a multi-agent Deep Reinforcement Learning (DRL) algorithm for Intelligent Reflecting Surface (IRS) and UAV-assisted wireless powered edge networks. Then, a lightweight scheduling algorithm was developed to solve the optimization problem.

In 2025, Tian *et al.* [3] developed a hybrid approach that combined Double Exponential Crayfish Optimization Algorithm (DECOA) with Deep Belief Networks (DBN). The DBN model was used to predict energy consumption, which reduced network energy usage by maintaining reliable connectivity.

In 2025, Ren et al. [4] employed K-means clustering to group Ground Users (GUs) into U clusters, and each group was assigned a serving UAV based on improved Particle Swarm Optimization (PSO). An EE maximization problem was then formulated by optimizing UAV 3D positions and transmit power of both GUs and UAVs.

In 2026, Hosseini et al. [5] developed an energy-efficient resource optimization framework. The proposed approach combined semantic-aware transmission, which reduced redundant data by extracting task-oriented information, with Non-Orthogonal Multiple Access (NOMA)-based spectrum sharing to enhance spectral efficiency and EE in dense medical environments.

2.2 Problem Definition

UAVs have become an important part of modern wireless communication networks because they quickly provide flexible and reliable connectivity to users and devices. Their ability to move easily and adapt to different environments improves communication coverage, network capacity, and service quality, especially in areas with high data traffic. However, the rapid growth of connected devices with limited battery power creates significant challenges in maintaining efficient network operation.

- ❖ Conventional methods suffer from limited onboard battery capacity, which restricts flight duration and continuous network operation. Efficient energy management becomes challenging due to the need for real-time coordination between UAVs and ground networks.
- ❖ Existing methods lack effective coordination among multiple UAVs, resulting in inefficient task scheduling and unnecessary energy consumption. Furthermore, most conventional approaches provide limited transparency in their decision-making process, which makes it difficult to understand the resource allocation and trajectory planning decisions.
- ❖ Prior methods struggle to adapt to rapidly changing network conditions. This leads to suboptimal energy utilization, inefficient UAV trajectory planning, and poor resource allocation. Reinforcement learning addresses these challenges by enabling UAVs to learn optimal decision-making policies through continuous interaction with the environment.
- ❖ Traditional models face challenges in scaling to large and dense deployments due to increasing computational complexity and communication costs. They also have limited adaptability to real-time network variations, which restricts their practical deployment in next-generation wireless communication systems.

3. System Model and Problem Formulation of UAV System with Static and Mobile Users

3.1 System Model

A wireless communication system [1] consisting of a set of static and mobile ground users g , which is considered within a predefined geographical area. The location of each ground user u at time t is represented by two-dimensional coordinates (a'_u, x'_u) . Due to network failures, natural disasters, or sudden increases in user traffic, the existing terrestrial communication infrastructure experiences service outages. To overcome these limitations and provide continuous wireless connectivity, a group Q of quad-rotor UAVs is deployed as ABS to serve ground users. The considered multi-UAV system enables UAVs to exchange state information with their neighboring UAVs, which improves coordination and network management. The position of each UAV $m \in Q$ at time t is represented by three-dimensional coordinates (a'_m, x'_m, c'_m) . Due to the high-altitude deployment of UAVs, a line-of-sight (LoS) communication link is assumed between UAVs and ground users to ensure reliable data transmission. Each ground user is connected to a single UAV that provides the strongest downlink signal based on Signal-to-Interference-plus-Noise-Ratio (SINR). The mathematical representation of SINR at time t is defined in Eq. (1).

$$\eta_{u,m}^t = \frac{\chi P (d_{u,m}^t)^{-\tau}}{\sum_{y \in g_{\text{int}}} \chi P (d_{u,y}^t)^{-\tau} + \sigma^2} \quad (1)$$

Here, τ implies path loss exponent, $\mathcal{X}$ indicates attenuation factor, σ^2 specifies Gaussian noise, $\mathcal{G}_{\text{int}}$ represents interfering UAVs, P implies transmit power, and $d_{u,m}^t$ denotes distance. To prevent collisions among UAVs, a minimum separation distance d_{coll} is maintained between any two UAVs. The mobility patterns of mobile ground users are modeled using widely adopted mobility models, including Random Waypoint (RWP), Random Walk (RW), and Gauss-Markov (GMM) models, which allow users to dynamically change their locations over time. Since ground users may be either stationary or mobile, UAVs are required to optimize their flight trajectories to provide continuous and reliable connectivity throughout the service area. Considering a communication channel with bandwidth B_v , the data transmission rate of each ground user is calculated based on Shannon's capacity equation, which is described in Eq. (2).

$$G_{u,m}^t = B_v \log_2(1 + \eta_{u,m}^t) \quad (2)$$

The considered UAV-assisted network is assumed to be an interference-limited system, where the coverage performance depends on SINR. Therefore, the connectivity performance of each UAV $m \in \mathcal{Q}$ at time t is computed in Eq. (3).

$$p_m^t = \sum_{u \in \mathcal{G}} v_m^t(u) \quad (3)$$

Here, $v_m^t(u) \in [0,1]$ indicates whether ground user u is connected to UAV m at time t .

3.2 Energy Consumption Model

During flight, each UAV $m \in \mathcal{A}$ consumes energy h_m^t at time t . The total energy consumption h_t consists of propulsion energy h_p and communication energy h_c , where $h_t = h_p + h_c$. Since communication energy is much smaller than propulsion energy, it is neglected in this work. Therefore, the propulsion energy of a rotary-wing UAV is considered. The closed-form analytical expression for propulsion power consumption of a rotary-wing UAV at time t is given in Eq. (4).

$$P(t) = \beta_0 \left(1 + \frac{3F^2}{S_{\text{tip}}^2} \right) + \beta_1 \left(\sqrt{1 + \frac{F^4}{4f_0^4}} + \frac{F^2}{2f_0^2} \right)^{\frac{1}{2}} + \frac{\beta_2}{2} F^3 \quad (4)$$

Here, β_0 , β_1 , and β_2 indicate constants; S_{tip} specifies the tip speed of the rotor blade, f_0 implies the mean hovering speed, and F denotes the speed of the UAV. In this model, the basic flight activities of the UAV, including hovering and acceleration, are considered when estimating energy consumption. The average propulsion power is calculated over all time steps as $\frac{1}{T} \sum_{t=1}^T P(t)$, and the total energy consumed by UAV m at time t is expressed in Eq. (5).

$$h_m^t = \gamma_t \cdot P(t) \quad (5)$$

Here, γ_t represents the time-step duration. The EE of UAV m is defined as the ratio of data throughput to energy consumed during time step t . It is expressed in Eq. (6).

$$\delta_m^t = \frac{\sum_{u \in \mathcal{K}} G_{u,m}^t}{h_m^t} \quad (6)$$

Increasing the number of connected users C_m^t allows UAV m to transmit more data $\sum_{u \in \mathcal{K}} G_{u,m}^t$ during time step t . For a given energy consumption h_m^t , higher data throughput results in improved EE α_{tot} . Therefore, the system's EE over all time steps is expressed in Eq. (7).

$$\alpha_{\text{tot}} = \frac{\sum_{t=1}^T \sum_{m \in \mathcal{S}} \sum_{u \in \mathcal{K}} G_{u,m}^t}{\sum_{t=1}^T \sum_{m \in \mathcal{S}} h_m^t} \quad (7)$$

3.3 Problem Formulation

The main objective is to maximize the EE of the system by optimizing the number of connected ground users, the UAV's three-dimensional trajectory, and energy consumed during service.

These optimizations are performed by ensuring that the UAVs operate within a fixed energy limit. Based on these requirements, the optimization problem is formulated as follows.

$$\max_{\forall m \in \mathcal{Z}, a_m^t, x_m^t, c_m^t, h_m^t, p_m^t} \alpha_{tot} \quad (8)$$

$$\text{s.t.} \quad \eta_{u,m}^t \geq \eta_{th}, \quad \forall v_m^t(u) \in [0,1], u, m, t \quad (9)$$

$$h_m^t \leq h_{\max}, \quad \forall m, t \quad (10)$$

$$a_{\min} \leq a_m^t \leq a_{\max}, \quad \forall m, t \quad (11)$$

$$x_{\min} \leq x_m^t \leq x_{\max}, \quad \forall m, t \quad (12)$$

$$c_{\min} \leq c_m^t \leq c_{\max}, \quad \forall m, t \quad (13)$$

Here, $a_{\min}$, $x_{\min}$, $c_{\min}$ and $a_{\max}$, $x_{\max}$, $c_{\max}$ imply the minimum and maximum coordinates of a , x , and c . The UAV's maximum budget is represented as $h_{\max}$. The constraints in Eq. (9)-(13) define the operating limits of UAVs. Eq. (9) ensures that all users satisfy the minimum SINR requirement. Eq. (10) limits the energy consumption of each UAV so that it remains within the available energy budget. Eqs. (11)-(13) restrict the UAVs to a specified operating area. When multiple UAVs or small-cell access points use the same frequency spectrum in a shared wireless environment, interference becomes more difficult to manage because it depends on the locations of UAVs. As the number of deployed UAVs increases, the optimization problem in Eq. (8) becomes more complex, and it is classified as an NP-hard problem. Multi-agent optimization becomes increasingly difficult when more than two UAVs operate in a shared environment, which makes conventional optimization methods ineffective. In dynamic networks, UAVs prioritize individual objectives over system EE due to changing network conditions. Multi-Agent Reinforcement Learning improves coordination and decision-making in these environments. Moreover, sharing information among neighboring UAVs increases communication overhead, but it enhances coordination and system performance. Therefore, a multi-agent reinforcement learning approach is adopted to maximize the system's EE.

4. Workflow Representation of Energy-Efficient UAV Trajectory and Resource Optimization using Reinforcement Learning

4.1 Step-wise Illustration of Proposed Model

UAVs are widely adopted for providing wireless connectivity to static and mobile ground users, especially in situations where terrestrial cellular infrastructure faces failures, overload conditions, or sudden increases in network demand. Due to their rapid deployment capability and flexible mobility, UAVs effectively improve network coverage and restore communication services in challenging environments. However, UAV-assisted networks face several issues, such as limited onboard energy resources and interference from neighboring UAV cells operating over the same frequency spectrum, which affect the EE of the system. Existing approaches mainly focus on optimizing UAV trajectories for static ground users while neglecting user mobility, dynamic network conditions, and interference effects among UAVs. Moreover, some methods assume an interference-free environment or depend on a centralized controller that requires continuous global information about user locations and network states for UAV decision-making. Such centralized solutions may not be practical in disaster scenarios due to high communication overhead, increased delays, and difficulties in accurately tracking user positions. To moderate these complexities, this paper implements the SAM-EMRL framework for optimizing EE in UAV-assisted wireless networks. The proposed method adopts a decentralized multi-agent reinforcement learning strategy, where each UAV independently learns optimal actions by collaborating with neighboring UAVs through local information exchange. The sparse attention mapping mechanism enables UAV agents to focus on the most relevant environmental information, including user mobility patterns, UAV states, interference levels, and energy availability, thereby improving learning efficiency and decision accuracy. In addition, the explainable learning capability of SAM-EMRL enhances transparency and reliability of UAV decision-making, making it more suitable for practical network deployments. By jointly optimizing UAV trajectories, user association, and resource allocation, the proposed model reduces energy consumption, improves normalized energy efficiency, and maintains reliable communication in dynamic network conditions. Therefore, SAM-EMRL provides an energy-efficient solution for intelligent management of UAV-assisted wireless networks. Fig. 1

depicts the network design of the designed approach for optimizing EE in UAV-assisted networks.

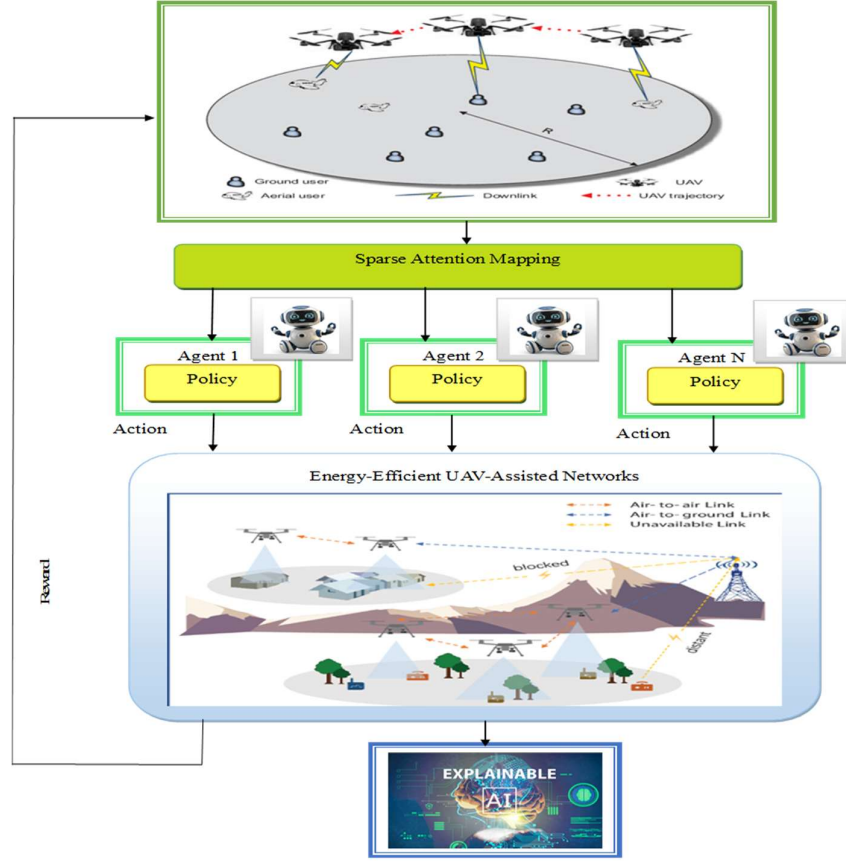


Fig. 1. Network Design of Developed Model for Optimizing EE in UAV-assisted Networks

4.2 Reinforcement Learning

RL [16] is a goal-driven learning approach in which an RL agent learns and improves its actions by interacting with an unknown and unmodeled environment. The agent aims to optimize its decisions to achieve the highest possible cumulative reward or the lowest cost function by following a self-learned control policy. The fundamental framework of RL is a Markov Decision Process (MDP), which consists of state space S , action space A , reward function R , and state transition probability P . At each time step t , the agent observes the current state $s_t \in S$, selects an action $a_t \in A$, receives an immediate reward $r_t(s_t, a_t)$, and moves to the next state based on the transition probability. The basic RL system includes a control policy and learning algorithm for updating the policy. The control policy determines the action to be taken according to the current state s_t of the environment, and it is represented by a function approximator. The learning algorithm updates the tunable parameters of the policy using observed state s_t and corresponding reward $r_t(s_t, a_t)$ to improve the agent's performance. The cumulative reward r_t reflects the control performance and guides the policy update process. Therefore, the primary objective of the learning algorithm is to obtain an optimal control policy that enables the agent to maximize the total cumulative reward r_t over time, and it is defined in Eq. (14).

$$R_t = \sum_{i=0}^{\infty} \alpha^i r_t(s_t, a_t) \quad (14)$$

In RL, a reward discount factor α ($0 < \alpha < 1$) is introduced to give higher priority to immediate rewards while reducing the influence of future rewards. The main goal of the RL agent is to determine an optimal control policy that achieves maximum cumulative reward throughout the learning process, which is described in Eq. (15).

$$\pi = \arg \max_{\pi} \left[\sum_{t=0}^{\infty} \alpha^t r_t(s_t, a_t) \right] \quad (15)$$

This reward serves as feedback to evaluate the agent's actions and guide the learning process toward an optimal policy.

4.3 Proposed SAM-EMRL for UAV Trajectory and Resource Optimization

The SAM-EMRL framework is designed to improve EE and decision interpretability in dynamic UAV-assisted wireless networks. However, conventional methods face challenges, including high computational complexity, excessive communication overhead among UAVs, and limited adaptability to dynamic network conditions. In addition, most existing deep reinforcement learning approaches function as black-box models, which makes it difficult to interpret the reasons behind UAV decision-making. To address these issues, the SAM-EMRL approach is proposed. This achieves efficient resource utilization, reduced energy consumption, and trustworthy autonomous decision-making in UAV-assisted wireless networks.

The sparse attention mapping mechanism selects the most relevant neighboring UAVs and network features, thereby reducing redundant information exchange and computational complexity. The multi-agent reinforcement learning enables cooperative decision-making for optimizing UAV trajectory, power allocation, and resource management to improve EE. Furthermore, the explainability module provides interpretable insights into factors that influence UAV decisions, enhancing transparency and reliability.

The first stage of the SAM-EMRL framework is the collection of real-time environmental information required for intelligent decision-making. Each UAV agent observes the current network conditions and gathers relevant information, including user communication requirements, channel characteristics, UAV energy status, interference conditions, and neighboring UAV information. The collected observations represent the current state of each UAV agent. Since UAV-assisted networks are highly dynamic, these states continuously change with user mobility, channel variations, and UAV movement. The observed information is then forwarded to the sparse attention mapping module for efficient processing.

Sparse Attention Mapping: The collected state information is processed using a sparse attention mapping mechanism [17] to identify the most influential relationships among UAV agents. In a conventional multi-agent reinforcement learning framework, every UAV exchanges information with all other UAVs, which increases communication overhead and computational complexity. To overcome this limitation, sparse attention mapping calculates attention weights among UAV agents based on their state relationships. These attention values indicate the importance of neighboring UAVs and environmental features that influence the current UAV's decision. The expression of sparse attention mapping is specified in Eq. (16).

$$U_r = \text{softmax} \left(\frac{qk^T}{\sqrt{j_k}} + \log(M) \right) + v^l \quad (16)$$

Here, q indicates the query matrix, k denotes the key matrix, v^l represents the value, j_k specifies the key dimension, M implies the sparse attention mapping matrix, and U_r represents sparse attention weights. Based on calculated attention scores, the most relevant UAV interactions and important features are selected. The selected information is then aggregated with local UAV state using attention weights to generate compact and informative feature representation. This sparse feature aggregation improves scalability and reduces redundant information processing.

MARL: The explainable attention-based representation is provided to the MARL module, where each UAV operates as an intelligent agent. Based on processed state information, the policy network generates optimal actions to improve network EE. The UAV agents learn cooperative strategies through continuous interaction with the environment. The generated actions include UAV trajectory adjustment, transmission power control, user association, spectrum allocation, and resource scheduling. During the training process, the MARL agents learn the relationship between their actions and network performance. Through repeated interactions, the agents gradually develop optimal policies that balance communication performance and energy consumption.

Energy-Efficient Action Execution and Reward Evaluation: After the action selection process, the UAV agents execute the chosen actions in the network environment. The performance of each action is evaluated using an energy-aware reward function.

The reward function evaluates the effectiveness of selected actions based on key indicators like EE, throughput, energy consumption, interference, and quality-of-service satisfaction. Actions that increase network EE receive higher rewards, which encourages UAV agents to select similar strategies in future decisions. Inefficient actions receive lower rewards, which allows the agents to adjust their future behavior.

Policy Update: The final stage of SAM-EMRL is the continuous learning and optimization process. The experience generated during interaction with the environment, including the current state, selected action, obtained reward, as well as the next state, is stored and used to update the reinforcement learning policy. The policy network parameters are optimized iteratively to improve the decision-making capability of UAV agents. As training progresses, the agents learn more effective strategies for trajectory control, power management, and resource allocation.

Explainable Attention-based Representation: The explainability module provides transparency in the decision-making process. This allows the SAM-EMRL framework to provide interpretable insights into why a specific action is selected. For instance, if a UAV changes its trajectory, the explanation mechanism indicates whether the decision was influenced by improving channel conditions, reducing energy consumption, avoiding interference, or satisfying user demand. Thus, the explainability module transforms the learned representation into an understandable decision interpretation. Fig. 2 shows the visual view of the SAM-EMRL approach for energy-efficient UAV-assisted networks.

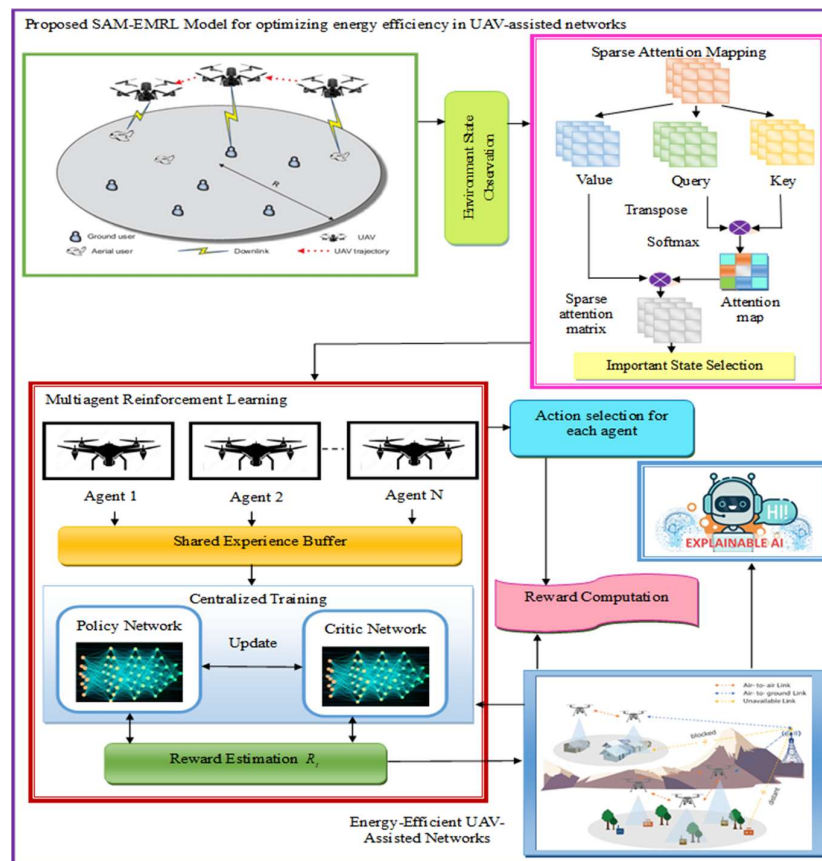


Fig. 2. Visual View of SAM-EMRL Approach for Energy-Efficient UAV-assisted Networks

5. Results and Discussion

5.1 Experimental setup

The proposed SAM-EMRL model was implemented in Python to optimize EE in UAV-assisted wireless networks. Its performance was evaluated through a series of experiments conducted under different network conditions and parameter settings using various performance metrics.

The proposed approach was compared with prior methods, like Multi-agent DRL [2], DECOA-based DBN [3], MAD-DDQN [18], and UTRP-DDPG [19], to measure its effectiveness. The experimental results show that the SAM-EMRL model achieves better EE and reliable communication, which makes it suitable for energy-efficient UAV-assisted wireless networks.

5.2 Performance analysis of the designed method using the number of UAVs

Fig. 3 depicts the performance of the proposed SAM-EMRL model by comparing it with existing methods, including Multi-agent DRL, DECOA-based DBN, MAD-DDQN, and UTRP-DDPG, based on the number of UAVs. The proposed model attains a higher fairness index, which demonstrates its ability to allocate network resources more evenly among ground users, thereby improving service fairness and reducing performance imbalance. It achieves greater total bits exchanged, which indicates efficient data transmission and higher network throughput. The model also supports a larger number of connected users, which highlights its capability to provide wider network coverage and maintain reliable connectivity for more users simultaneously. Furthermore, it obtains higher normalized EE, which reflects the effective utilization of available energy while maintaining high communication performance. In addition, the proposed SAM-EMRL model records lower total energy consumption, which indicates its ability to minimize energy usage through efficient UAV trajectory optimization and resource management. The SAM-EMRL model achieves a better number of connected users of 11.26% than Multi-agent DRL, 17.44% than DECOA-based DBN, 15.07% than MAD-DDQN, and 3.69% than UTRP-DDPG at number of UAVs=12. These outcomes show that the developed model provides a balanced and energy-efficient solution for UAV-assisted wireless networks by improving communication performance and reducing energy usage.

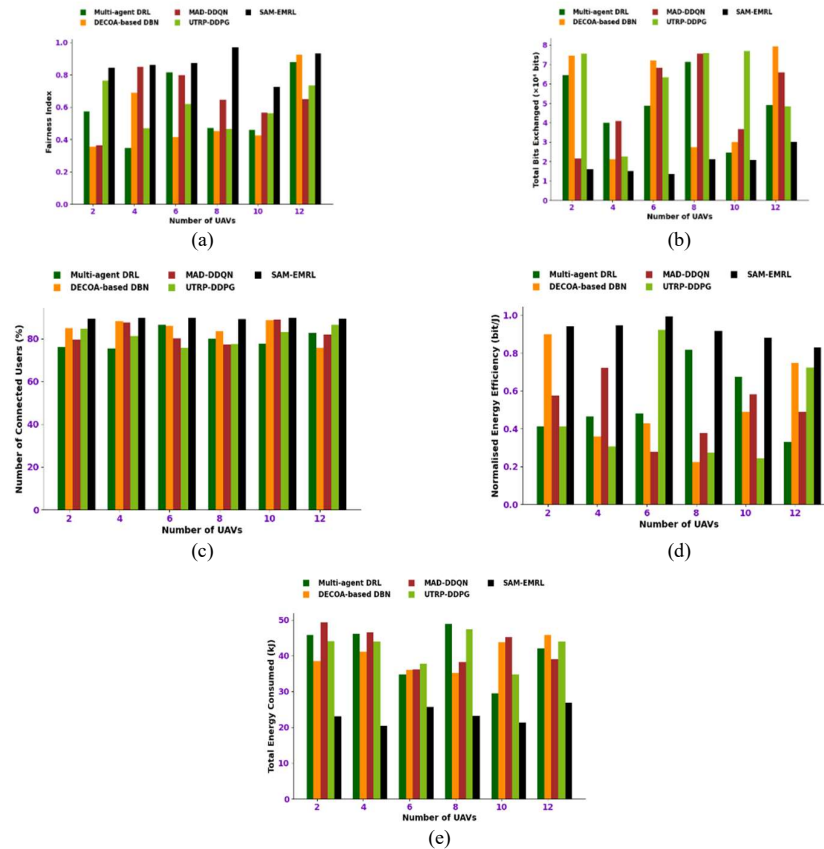


Fig. 3. Number of UAVs based Efficiency Assessment of Developed Method with respect to (a) Fairness Index (b) Total Bits Exchanged (c) Number of Connected Users (d) Normalized Energy Efficiency (e) Total Energy Consumed

5.3 Episode-based Efficiency Analysis of Proposed Model

The performance of the SAM-EMRL approach is analyzed by varying the number of episodes from 50 to 250, as shown in Fig. 4. The analysis is carried out using key metrics, including total

energy consumed, normalized energy efficiency, fairness index, and number of connected users. As the number of training episodes increases, the model gradually learns an effective resource allocation and UAV trajectory optimization strategy that results in higher normalized EE, which indicates improved utilization of available energy resources. At the same time, the total energy consumed decreases, which demonstrates the model's ability to reduce unnecessary energy expenditure through better decision-making. The fairness index also improves with increasing episodes, which reflects a more balanced distribution of network resources among ground users. In addition, the number of connected users increases, which shows that the proposed model provides wider network coverage and maintains reliable communication for more users. The SAM-EMRL approach attains a fairness index of 0.99236 at episodes=200. Moreover, the results indicate that the SAM-EMRL model achieves stable convergence and consistently improves network performance. Table 1 displays the learning rate-based performance evaluation of the designed SAM-EMRL approach. The developed SAM-EMRL method attains a higher fairness index of 53.97% than Multi-agent DRL, 56.97% than DECOA-based DBN, 43.53% than MAD-DDQN, and 59.37% than UTRP-DDPG at learning rate=0.011.

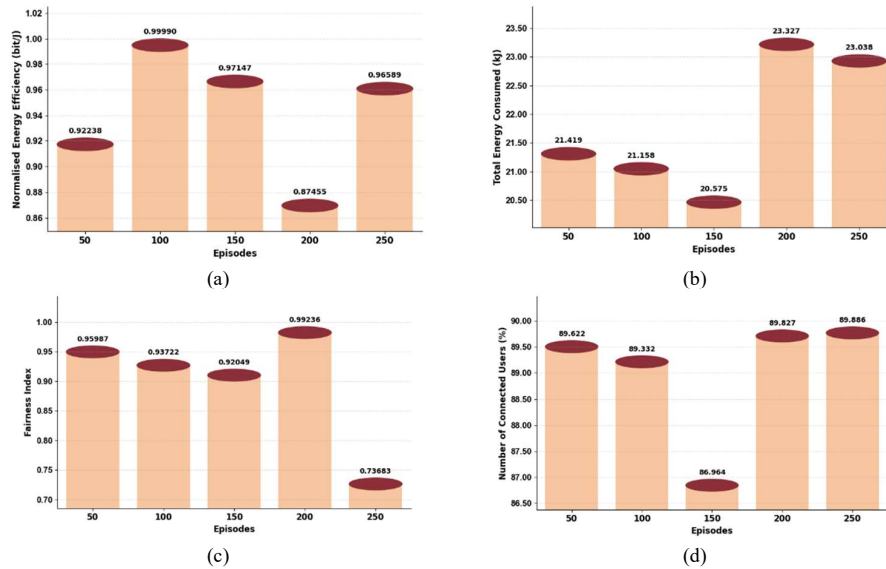


Fig. 4. Efficacy Evaluation of SAM-EMRL Model based on Episodes concerning (a) Normalized Energy Efficiency (b) Total Energy Consumed (c) Fairness Index (d) Number of Connected Users

Table 1. Learning rate-based performance assessment of developed method

Rate	Learning	Multi-agent DRL [2]	DECOA-based DBN [3]	MAD-DDQN [18]	UTRP-DDPG [19]	SAM-EMRL
Normalised Energy Efficiency						
0.011		0.592446	0.5927	0.504327	0.841418	0.992207
0.022		0.423604	0.465821	0.526713	0.23421	0.784582
0.033		0.479954	0.772156	0.430388	0.619946	0.991219
0.044		0.941575	0.389255	0.206998	0.926192	0.976326
0.055		0.616771	0.692609	0.333471	0.873923	0.909325
Number of Connected Users						
0.011		80.93493	89.32838	79.79667	82.29061	89.41473
0.022		78.53313	75.08792	81.35529	86.62843	89.97622
0.033		86.98705	84.26939	79.17664	80.90407	88.08393
0.044		75.59033	86.30952	85.37729	78.09652	89.80741
0.055		81.07509	84.37225	79.51582	81.41492	89.45831
Total Energy Consumed						
0.011		42.1165	32.20368	21.78059	22.60862	21.43749
0.022		27.05186	25.18805	35.06967	37.06233	21.73116
0.033		40.66215	47.52671	48.40156	43.97306	38.61351
0.044		25.95606	49.60179	35.44105	42.51817	25.93954
0.055		34.45916	34.99674	29.55325	27.79292	21.77512
Fairness Index						
0.011		0.453587	0.430233	0.556461	0.400392	0.985417
0.022		0.839091	0.341303	0.574238	0.815003	0.865609
0.033		0.805052	0.847383	0.75518	0.318889	0.897515
0.044		0.595286	0.917372	0.546747	0.394523	0.959332
0.055		0.640769	0.523749	0.562292	0.911646	0.999041

5.4 Performance Evaluation of SAM-EMRL Model using Mobility Models

Fig. 5 shows the effectiveness of the designed SAM-EMRL model by comparing it with existing approaches using different mobility models, including Static, GMM, RW, and RWP. The SAM-EMRL model achieves the lowest total energy consumption across all mobility models, which demonstrates its ability to optimize UAV trajectory planning and resource allocation while minimizing unnecessary energy expenditure. It attains the highest normalized energy efficiency, which indicates effective utilization of available energy resources to maintain reliable communication. The proposed model records the highest fairness index, thereby ensuring balanced service quality and reducing resource allocation disparity. Furthermore, SAM-EMRL consistently supports the largest number of connected users, which reflects its ability to provide wider network coverage and maintain stable connectivity in different mobility scenarios. These results show that the proposed approach effectively optimizes EE and communication performance in UAV-assisted wireless networks operating in diverse user mobility conditions.

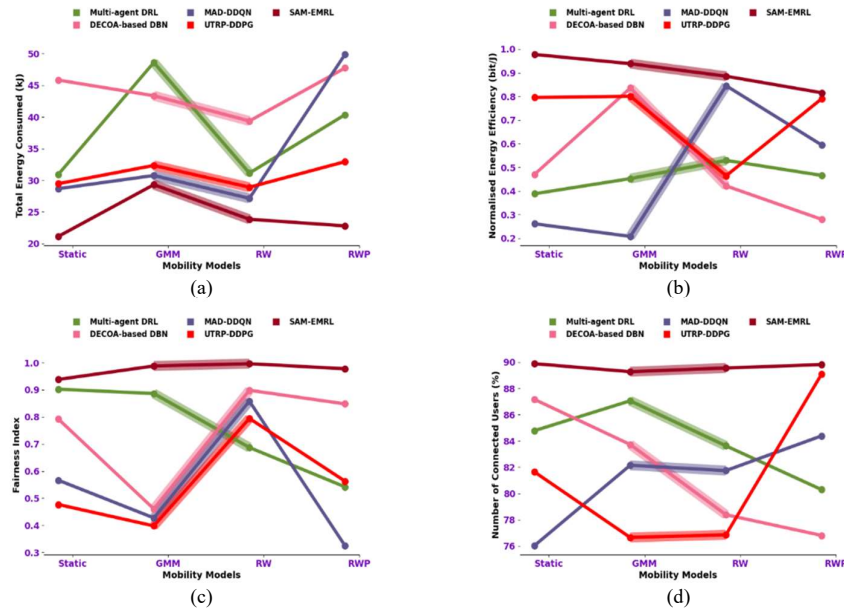


Fig. 5. Efficiency Assessment of SAM-EMRL Method using Mobility Models in terms of (a) Total Energy Consumed (b) Normalized Energy Efficiency (c) Fairness Index (d) Number of Connected Users

5.5 Reinforcement Learning-based Efficiency Analysis of Developed Model

Fig. 6 portrays the reinforcement learning-based performance of the proposed SAM-EMRL model using training metrics of reward and penalty over multiple episodes. The SAM-EMRL model achieves the highest cumulative reward throughout the training process. The reward increases rapidly during initial episodes and converges quickly to a stable value of approximately 418, which demonstrates efficient learning and faster convergence. However, Multi-agent DRL, DECOA-based DBN, MAD-DDQN, and UTRP-DDPG converge to lower reward values, which indicates less effective policy learning. The higher reward obtained by SAM-EMRL reflects its ability to identify optimal UAV trajectory planning and resource allocation strategies that maximize long-term network performance. Moreover, the SAM-EMRL model rapidly reduces the penalty and stabilizes at the lowest value of nearly 1.0. In contrast, the existing methods converge to higher penalty values, which shows slower adaptation and less efficient decision-making. The lower penalty achieved by SAM-EMRL demonstrates its capability to avoid suboptimal actions and improve the quality of the learned policy. Furthermore, the proposed model exhibits higher cumulative reward and lower penalty than benchmark methods, which highlights its superior capability for optimizing EE in UAV-assisted wireless networks.

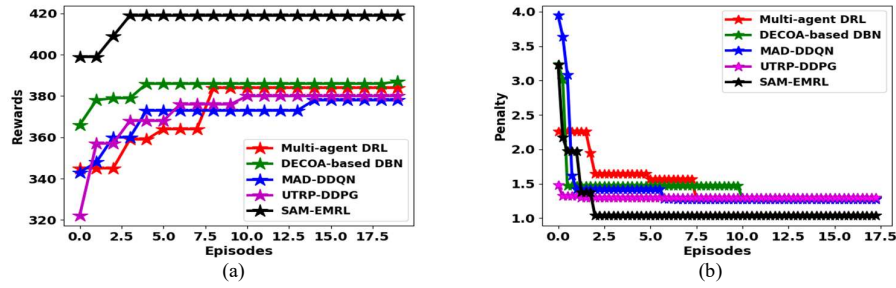


Fig. 6. Reinforcement Learning-based Efficacy Assessment of SAM-EMRL Model based on (a) Rewards (b) Penalty

5.6 Explainability Result of Developed SAM-EMRL Model

The explainability results help to understand how the UAV makes energy-efficient decisions during network operation. This allows network operators to verify that the selected actions improve EE by maintaining reliable communication. As a result, the explainability module increases the transparency, trust, and practical usability of UAV-assisted networks. Fig. 7 depicts the explainability result of the proposed SAM-EMRL model.

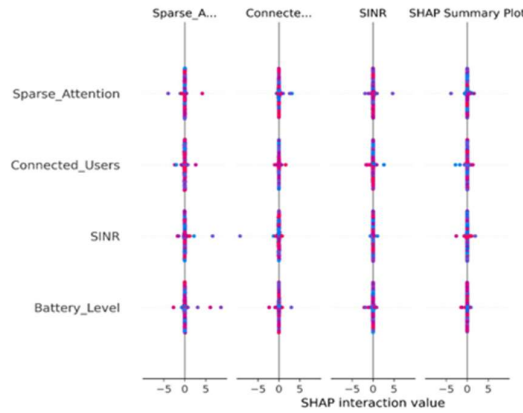


Fig. 7. Explainability Outcome of SAM-EMRL Model

6. Conclusion

This work proposed a SAM-EMRL framework to optimize EE in UAV-assisted wireless networks by combining sparse attention mapping with explainable MARL for efficient resource allocation, cooperative decision-making, and trajectory optimization. The proposed method reduces computational complexity by eliminating unnecessary attention computations while enhancing interpretability of agent decisions that lead to improved EE, network performance, and resource utilization compared with classical approaches. However, the proposed model has limitations, including dependence on sparse attention patterns, additional computational overhead introduced by the explainability component, and performance degradation in highly dynamic environments with varying network conditions. Future work will focus on developing adaptive sparse attention mapping strategies, lightweight explainability mechanisms, and implementing federated multi-agent learning approaches to improve robustness and efficiency of the proposed framework.

References

- [1] Omoniwa, B., Galkin, B. and Dusparic, I., "Communication-enabled deep reinforcement learning to optimise energy-efficiency in UAV-assisted networks," *Vehicular Communications*, vol. 43, pp. 100640, 2023.

- [2] Wang, X., Li, J., Wu, J., Guo, L. and Ning, Z., "Energy efficiency optimization of IRS and UAV-assisted wireless powered edge networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 18, no. 7, pp. 1297-1310, 2024.
- [3] Tian, Y., Khan, A., Ahmad, S., Mohsan, S.A.H., Karim, F.K., Hayat, B. and Mostafa, S.M., "An Optimized Deep Learning Framework for Energy-Efficient Resource Allocation in UAV-Assisted Wireless Networks," *IEEE Access*, 2025.
- [4] 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.
- [5] Hosseini, M., Hosseini, P., Abouei, J., Baladi, E. and Mohammadi, A., "Energy-Efficient Resource Optimization for UAV-Assisted Semantic-Aware IoMT Networks with NOMA," *Physical Communication*, pp. 103203, 2026.
- [6] Liu, Q., Zhou, X., Yang, H., Zhang, J. and Chen, J., "Deep Reinforcement Learning-Based Cross-Layer Energy Efficiency Optimization in UAV-MEC Systems," *Mobile Networks and Applications*, vol. 30, no. 5, pp. 817-830, 2025.
- [7] Pasha, S.A. and Vali Mohamad, N.M., "Energy efficient PSO integrated TD3 reinforcement learning with nonlinear energy harvesting in UAV assisted D2D networks," *International Journal of Networked and Distributed Computing*, vol. 14, no. 1, pp. 12, 2026.
- [8] Salim, M.M., Rabie, K.M. and Muqaibel, A.H., "Energy-efficient irregular RIS-aided UAV-assisted optimization: A deep reinforcement learning approach," *IEEE Internet of Things Journal*, 2025.
- [9] Sun, G., Zhang, L., Li, J., Wu, J., Wang, J., Sun, Z., Zhao, C. and Leung, V.C., "Age of information optimization in laser-charged uav-assisted iot networks: A multi-agent deep reinforcement learning method," *IEEE Transactions on Network Science and Engineering*, 2025.
- [10] Ahmad, W., Islam, U., Abdulkadhem, A.A., Shah, B., Moreira, F. and Abbas, A., "Enhancing reliable and energy-efficient UAV communications with RIS and deep reinforcement learning," *PeerJ Computer Science*, vol. 11, pp. e3031, 2025.
- [11] Chen, Y., Kang, H., Li, J., Sun, G., Wang, B., Wang, J., Liang, C., Liang, S. and Niyato, D., "Joint resource management for energy-efficient UAV-assisted SWIPT-MEC: A deep reinforcement learning approach," *IEEE Internet of Things Journal*, 2025.
- [12] Umar, A.R., Albinsaid, H., Wei, C.P. and Li, C.P., "Deep reinforcement learning for RIS-assisted multi-UAV MU-MISO communication networks: Sum-rate and energy efficiency maximization," *IEEE Open Journal of Vehicular Technology*, 2025.
- [13] Yi, M., Wang, X., Liu, J., Zhang, Y. and Hou, R., "Meta-reinforcement learning for timely and energy-efficient data collection in solar-powered uav-assisted iot networks," *IEEE Transactions on Communications*, 2025.
- [14] 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.
- [15] Li, J., Cui, J., Huang, C., Zhu, J., Chu, Z., Chen, M., Xiao, P. and Tafazolli, R., "Hybrid Bit and Semantic Communications for UAV-Enabled Wireless Power Transfer Networks: A Decision-Assisted Deep Reinforcement Learning Approach," *IEEE Journal on Selected Areas in Communications*, 2026.
- [16] Liu, B., Zhao, D., Lu, X. and Liu, Y., "Process control in semiconductor manufacturing based on deep distributional soft actor-critic reinforcement learning," *IEEE Transactions on Semiconductor Manufacturing*, vol. 38, no. 2, pp. 210-231, 2025.
- [17] Zhang, D., Yan, J., Polat, K., Alhudhaif, A. and Li, J., "Multimodal joint prediction of traffic spatial-temporal data with graph sparse attention mechanism and bidirectional temporal convolutional network," *Advanced Engineering Informatics*, vol. 62, pp. 102533, 2024.
- [18] Omoniwa, B., Galkin, B. and Dusparic, I., "Optimizing energy efficiency in UAV-assisted networks using deep reinforcement learning," *IEEE Wireless Communications Letters*, vol. 11, no. 8, pp. 1590-1594, 2022.
- [19] Luo, C., Zhang, J., Guo, J., Hong, Y., Chen, Z. and Gu, S., "Energy efficiency maximization in RISs-assisted UAVs-based edge computing network using deep reinforcement learning," *Big Data Mining and Analytics*, vol. 7, no. 4, pp. 1065-1083, 2024.