SCIENTIFIC OASIS

Decision Making: Applications in Management and Engineering

Journal homepage: www.dmame-journal.org
ISSN: 2560-6018, eISSN: 2620-0104

Autonomous Resource Allocation for QoS Enhancement in Next-Generation Heterogeneous Wireless Systems

Jafaar Fahad A. Rida[1,*] Elaf Noori Saddam[2], Basim Abood[3],

1 Department of Communication Engineering, College of Engineering, University of Sumer, Dhiqar, Iraq
2 Communication Engineering, College of Engineering, University of Sumer, Dhiqar, Iraq.
3 Communication Engineering, College of Engineering, University of Sumer, Dhiqar, Iraq.

ARTICLE INFO

Article history:
Received 12 March 2026
Received in revised form 28 April 2026
Accepted 17 June 2026
Available online 30 June 2026

Keywords:
Mobile Wireless Networks,
Heterogeneous Networks (HetNet),
Deep Reinforcement Learning (DRL),
Autonomous Resource Allocation,
Decision Support System (DSS)

ABSTRACT

The increasing heterogeneity of Beyond Fifth-Generation (B5G) and Sixth-Generation (6G) wireless networks, together with rapidly changing traffic demands, fluctuating channel conditions, and diverse Quality-of-Service (QoS) requirements, presents substantial challenges for autonomous resource allocation. Conventional static and reactive approaches lack the flexibility required for such dynamic environments, while existing learning-based techniques generally provide limited support for predictive, multi-objective, and temporally stable resource decisions. To address these limitations, this paper proposes a hybrid Decision Support System (DSS) that unifies Multi-Objective Optimization (MOOP), Deep Reinforcement Learning (DRL), a Predictive Environment Model (PEM), and Stability-Aware Optimization (SAO) within an autonomous resource-allocation framework. By combining real-time network observations with forecasts of forthcoming environmental conditions, the DSS autonomously determines resource allocations across multiple, potentially conflicting objectives. Within the framework, DRL facilitates adaptive policy learning, PEM anticipates future network states to support proactive decision-making, MOOP manages trade-offs among throughput, latency, fairness, and energy efficiency, and SAO promotes temporal consistency across consecutive allocation decisions. Communication-link quality is further assessed through Signal-to-Interference-plus-Noise Ratio (SINR) and data rate. Extensive simulation experiments compare the proposed DSS against static, reactive, and conventional DRL-based resource-allocation approaches. The findings demonstrate improvements in communication efficiency, resource utilisation, latency, fairness, and energy efficiency. Convergence and temporal evaluations further reveal faster and more consistent decision behaviour, while SINR and data-rate analyses indicate enhanced communication-link quality under the assumed operational conditions. Overall, the findings demonstrate that integrating adaptive learning, predictive environmental awareness, multi-objective optimisation, and stability-aware decision-making within a DSS provides an effective approach to autonomous resource management in dynamic B5G and 6G

* *Corresponding author.*
E-mail address: jafaarfahad@gmail.com

https://zenodo.org/records/24197529

heterogeneous wireless networks while maintaining balanced trade-offs among competing objectives.

1. Introduction

Fifth-generation (5G), B5G, and 6G wireless networks have created increasingly heterogeneous and dynamic communication environments [1]. The coexistence of macro base stations, small cells, heterogeneous users, diverse service demands, and time-varying wireless conditions introduces substantial challenges for autonomous network resource management. Resource allocation in such environments must simultaneously account for spectrum availability, transmission power, user association, interference, traffic dynamics, and QoS requirements. Conventional resource-allocation approaches typically rely on predefined rules, static optimisation, or demand-driven control mechanisms [2]. Although these approaches may perform adequately under relatively stable network conditions, their effectiveness can deteriorate when traffic loads, channel states, interference levels, and user mobility vary continuously. Autonomous resource management therefore requires decision mechanisms capable of observing the current network state, anticipating future conditions, evaluating competing alternatives, and selecting appropriate resource-allocation actions [3].

Within heterogeneous wireless networks, these requirements naturally constitute a multi-criteria decision-making problem. Network controllers must determine how available resources should be distributed among competing users while balancing throughput, latency, fairness, and energy efficiency. These objectives inherently involve trade-offs; for example, increasing transmission power may improve throughput while simultaneously raising energy consumption and interference. Consequently, an effective autonomous resource-allocation mechanism should not optimise individual objectives independently but should coordinate decisions across multiple QoS and operational criteria [4].

The central research problem addressed in this study concerns autonomous resource-allocation decision-making in heterogeneous wireless networks characterised by continuously evolving network conditions and service requirements. Resource-allocation actions must be selected under power, spectrum, interference, and QoS constraints while simultaneously considering multiple, potentially conflicting performance objectives. The network controller continuously observes the heterogeneous wireless environment and determines allocation actions involving transmission power, bandwidth allocation, user association, and scheduling. These decisions should enhance overall network utility while satisfying QoS requirements and preventing unstable resource-allocation behaviour [32]. The challenge therefore extends beyond identifying an optimal allocation at a particular instant; it requires adaptive decision-making over time under uncertain and changing conditions.

Accordingly, this study formulates autonomous wireless resource allocation as an intelligent multi-objective decision problem in which throughput, latency, fairness, and energy efficiency constitute the principal decision criteria. DRL, PEM, MOOP, and SAO are incorporated into the decision process. From a decision-making perspective, the proposed framework operates as an Intelligent Decision Support System (IDSS) for autonomous wireless network management. The network controller functions as an independent decision-maker that continuously observes network conditions, anticipates future environmental states, evaluates multiple performance criteria, and selects resource-allocation actions. Its decision variables encompass transmission power, bandwidth allocation, user association, and scheduling, while resource availability, QoS, interference, and stability define the principal constraints governing these decisions.

The framework consequently supports decision-making by integrating current network observations, predicted future conditions, and multi-objective optimisation before determining the final resource-allocation action. Unlike conventional optimisation methods that primarily derive allocations from the current network state, the proposed approach embeds prediction and adaptive

learning directly within the decision process. PEM anticipates potential future network conditions, DRL learns an adaptive allocation policy, MOOP captures trade-offs among competing objectives, and SAO regulates temporal variations across successive decisions. Through this integration, the framework facilitates autonomous and context-aware resource allocation in dynamic heterogeneous wireless networks, as illustrated in Figure 1 [5].

Figure 1 presents a multi-tier heterogeneous network (HetNet) architecture consisting of macro cells and small cells that collectively provide wireless coverage within a dense user environment. Network users exhibit heterogeneous QoS requirements associated with enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and massive Machine-Type Communications (mMTC). Macro cells provide wide-area coverage and mobility support, whereas small cells enhance capacity and spectral efficiency in high-density areas. The network operates under dynamic and non-stationary conditions in which user distributions, traffic demands, and channel states continuously evolve, thereby necessitating intelligent and adaptive resource-allocation strategies.

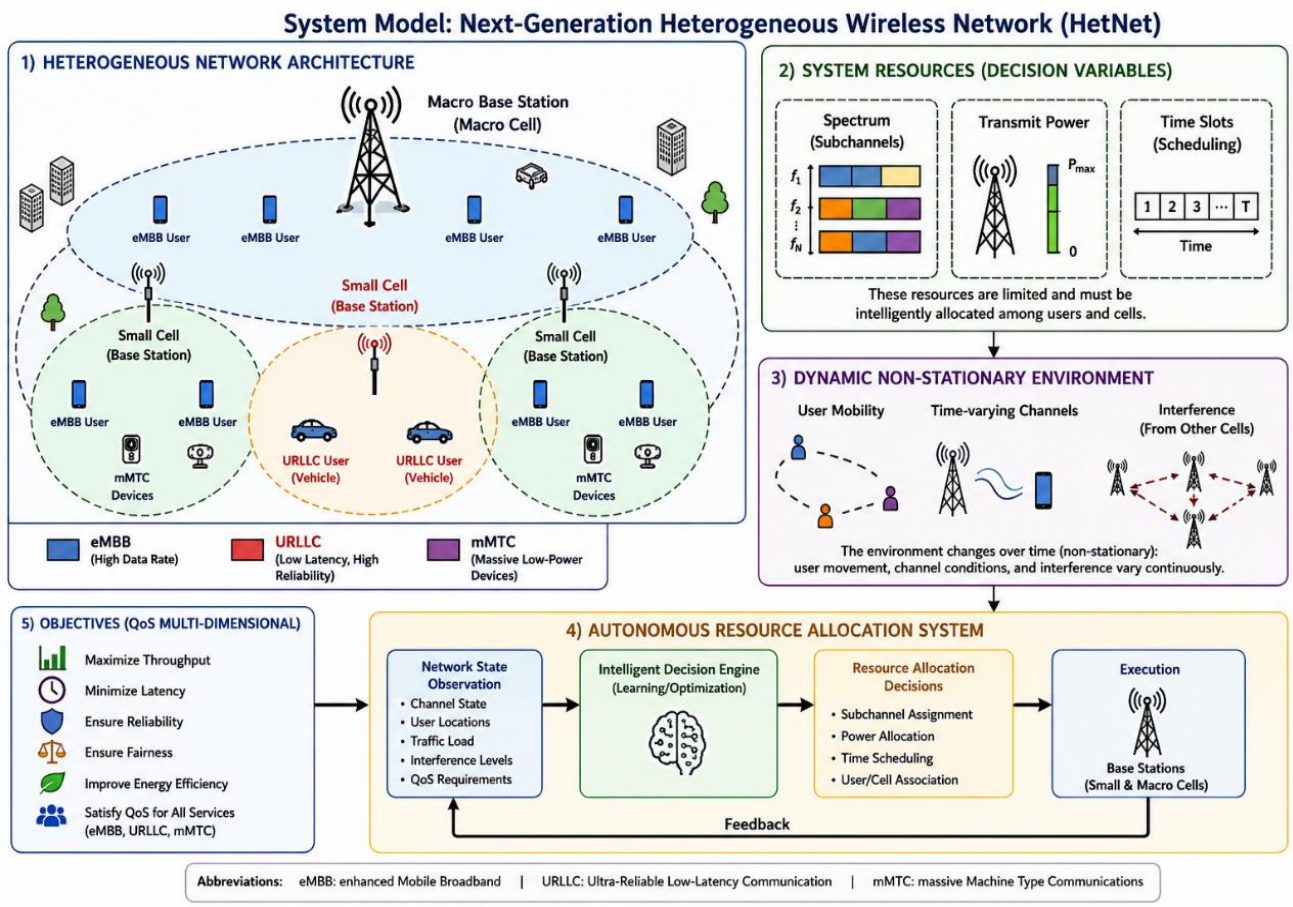


Fig.1: Architecture of a Multi-Layer Heterogeneous Wireless Network with Heterogeneous QoS Requirements

Despite substantial advances in AI-assisted resource allocation for heterogeneous wireless networks, developing and deploying a fully autonomous and intelligent resource-management framework capable of responding to rapidly changing network conditions in real time while preserving system stability and scalability remains a major challenge [10]. Existing approaches, particularly those employing DRL and optimisation methods, remain insufficient for simultaneously satisfying multidimensional QoS requirements in highly non-stationary and interference-intensive environments. Many existing models are developed under simplified or partially stationary conditions and consequently lack the robustness required for realistic environments characterised by bursty traffic and

mobile users. Furthermore, fast adaptive decision-making has not yet been adequately integrated with mechanisms that ensure long-term stability, creating an unresolved trade-off between responsiveness and reliability. These limitations highlight the need for a unified framework that coherently combines real-time adaptability, scalable architectures, and QoS assurance across varying network conditions, which remains an open and challenging problem for next-generation wireless systems [11; 12].

Existing research has established the effectiveness of DRL, multi-objective optimisation, predictive modelling, and QoS-aware resource-allocation techniques in addressing specific aspects of wireless network management. Nevertheless, important limitations emerge when these approaches are considered within an integrated decision-making framework. Many existing resource-allocation methods optimise only one or a limited set of network performance indicators, potentially creating undesirable trade-offs among throughput, latency, fairness, and energy efficiency. Conventional DRL-based allocation approaches also generally derive decisions from the observed network state without explicitly accounting for anticipated environmental conditions [6].

Although predictive resource-management methods can facilitate proactive decision-making, they may not adequately preserve consistency across successive allocation decisions. Similarly, multi-objective approaches provide mechanisms for reconciling competing criteria, but their integration with adaptive learning and predictive environmental information remains insufficiently explored, as illustrated in Figure 1. Collectively, these limitations demonstrate the need for an integrated decision-support mechanism that unifies adaptive learning, environmental prediction, multi-objective decision criteria, and temporal stability within a single resource-allocation process.

Accordingly, this study proposes a hybrid framework integrating DRL, PEM, MOOP, and SAO to support autonomous resource-allocation decisions in next-generation heterogeneous wireless systems. The principal objective is to develop an autonomous intelligent decision-support framework for multi-objective resource allocation in next-generation heterogeneous wireless networks [7]. Specifically, the study formulates wireless resource allocation as a multi-criteria decision-making problem involving throughput, latency, fairness, and energy efficiency; develops a DRL-based autonomous decision mechanism capable of adapting allocation decisions to dynamic network conditions; incorporates PEM to anticipate future network states and facilitate proactive allocation; applies MOOP to reconcile conflicting QoS and energy-related criteria; and integrates SAO to minimise undesirable variations between consecutive allocation decisions. The proposed framework is further evaluated against conventional resource-allocation and DRL-based approaches across different wireless network environments. In addition, the study examines the practical and managerial implications of autonomous resource allocation for network operators and wireless network managers.

1.1 Decision-Oriented Formulation

The study conceptualises autonomous wireless resource allocation as an intelligent multi-criteria decision-making problem in which throughput, latency, fairness, and energy efficiency constitute the principal decision criteria, subject to network and QoS constraints.

1.2 Autonomous Resource Allocation Decision Support Framework

The proposed hybrid architecture integrates DRL, PEM, MOOP, and SAO to support autonomous resource-allocation decisions within heterogeneous wireless networks.

1.3 Predictive Decision Capability

PEM is embedded within the decision process to provide information regarding anticipated network conditions, enabling the DRL-based controller to make proactive rather than exclusively reactive resource-allocation decisions.

1.4 Multi-Criteria Resource Allocation

The framework simultaneously considers throughput, latency, fairness, and energy efficiency, thereby explicitly addressing the conflicting objectives inherent in autonomous network resource management.

1.5 Stability-Aware Decision Mechanism (Sao)

SAO regulates temporal variations across successive resource-allocation decisions, thereby reducing instability within the autonomous decision process.

1.6 Comprehensive Evaluation

The proposed framework is evaluated against conventional allocation methods and DRL-based baselines using multiple performance measures, including prediction accuracy, convergence, SINR, data rate, latency, fairness, and energy efficiency.

1.7 Engineering Management Perspective

The study positions autonomous resource allocation as a decision-support mechanism through which telecom operators and network managers can make adaptive resource-management decisions under dynamically changing network conditions [8; 10].

A MOOP is formulated to address resource allocation and system optimisation within the considered heterogeneous network (HetNet). The formulation reflects the competing performance criteria that must be jointly optimised in next-generation wireless networks [15]. In particular, the framework seeks to maximise system throughput to improve overall spectral efficiency and accommodate high-data-rate services such as enhanced Mobile Broadband (eMBB). It simultaneously minimises end-to-end latency to satisfy stringent delay requirements, particularly for Ultra-Reliable Low-Latency Communications (URLLC), while improving energy efficiency by reducing power consumption at base stations and user equipment. Fairness is also incorporated to prevent resource starvation and promote equitable distribution among different user groups [16].

Furthermore, the formulation ensures compliance with service-specific QoS requirements by allocating resources according to the performance needs of eMBB, URLLC, and massive Machine-Type Communications (mMTC). Rather than treating these objectives independently, the mathematical formulation evaluates them jointly, thereby enabling explicit analysis of conflicting trade-offs such as throughput versus energy efficiency and latency versus fairness. Practical wireless-network constraints are also incorporated, including spectrum scarcity, spectrum reuse and associated interference in dense heterogeneous deployments, and service-specific QoS requirements such as minimum data rates, maximum latency, and reliability thresholds. The resulting MOOP can be addressed using advanced solution techniques, including machine-learning and evolutionary optimisation methods, to determine resource allocations that balance performance objectives against physical network constraints [11; 12].

The proposed framework has implications for multiple stakeholders involved in wireless network operation and management. Telecom operators can employ the decision-support mechanism to facilitate dynamic resource utilisation and QoS management, while network managers can use it to reconcile competing operational priorities involving capacity, latency, fairness, and energy consumption. Predictive information can additionally assist infrastructure planners in assessing future resource demand and network conditions. From an engineering-management perspective, the framework provides an autonomous mechanism for supporting operational decisions in heterogeneous wireless networks, reducing dependence on manually configured or static allocation policies. Section 9 concludes the study [13; 14].

DRL has emerged as an important mechanism for autonomous resource management in

heterogeneous and next-generation wireless networks because allocation policies can be learned through continuous interaction with dynamic environments. A comprehensive survey of DRL-based radio resource allocation and management highlighted its capacity to address complex resource-management problems without requiring an explicit analytical solution for every possible network condition [1]. However, as a survey, that work did not develop an integrated allocation mechanism combining prediction, multi-objective optimisation, and stability-aware control. The potential of DRL for energy-efficient resource management has also been demonstrated through an energy-efficient rate-splitting multiple-access framework that jointly considered communication performance and energy consumption [5]. Nevertheless, that investigation focused on a specific multiple-access scenario rather than a general heterogeneous network involving simultaneously conflicting QoS criteria.

The applicability of DRL to data-transmission optimisation in 6G software-defined networks has likewise been demonstrated, supporting its role in adaptive resource management within emerging 6G environments [22]. However, the emphasis remained primarily on transmission optimisation without explicitly integrating PEM with a stability-aware multi-objective decision mechanism. Multi-agent DRL has also been applied to efficient end–edge–cloud task offloading in 6G networks, demonstrating the value of distributed learning for resource-related decisions in dynamic environments [28]. Task offloading, however, constitutes a more specialised computational resource-management problem, whereas the present study simultaneously addresses wireless resource-allocation criteria involving throughput, latency, fairness, and energy efficiency.

QoS-aware adaptive resource management has further been investigated through intelligent congestion control for 5G URLLC software-defined networks using a reinforced duelling deep Q-network [31]. Although particularly relevant to latency-sensitive resource management, that approach concentrates primarily on congestion control and URLLC-oriented allocation rather than providing a unified multi-objective framework incorporating prediction, adaptive learning, and temporal stability. Dynamic scheduling has similarly been addressed through EdgeSched-DQN for edge-cloud environments, further demonstrating the usefulness of DQN-based learning for adaptive scheduling [35].

Nevertheless, its application domain concerns edge-cloud task scheduling rather than heterogeneous wireless resource allocation, and it does not explicitly optimise wireless throughput, latency, fairness, and energy efficiency simultaneously. Collectively, these studies establish the ability of DRL to support adaptive decision-making across dynamic wireless and edge environments. However, the approaches represented within the current literature generally address specialised resource-management problems or prioritise selected performance objectives. This limitation motivates a more integrated decision framework capable of jointly incorporating multiple QoS criteria, anticipated network conditions, and temporal stability.

Wireless resource allocation is inherently multi-objective because network operators must reconcile several competing performance criteria. Multi-objective optimisation has been investigated for active intelligent reflecting surface (IRS)-aided multi-group multicast systems incorporating energy harvesting and integrated sensing and communication, demonstrating the importance of jointly considering multiple system objectives in emerging wireless architectures [18]. However, that work addresses an active IRS-assisted multicast setting rather than autonomous resource management within heterogeneous cellular networks. A multi-objective evolutionary approach has also been employed to optimise energy savings and coverage in edge WiFi sensor deployment, illustrating the capacity of evolutionary optimisation to reconcile conflicting objectives [4]. Nevertheless, the underlying problem is deployment-oriented and does not incorporate adaptive DRL-based decision-making or predictive network-state information.

The significance of multi-objective decision-making has additionally been demonstrated through service-based application scheduling across multi-cluster cloud-edge-IoT infrastructures [7]. Although this approach highlights the need to balance competing objectives within complex resource-scheduling

environments, its focus lies primarily on cloud-edge-IoT service scheduling rather than wireless resource allocation governed by channel, interference, and QoS constraints. Multi-index evaluation learning has similarly been applied to computation-offloading optimisation for the power Internet of Things, demonstrating the importance of considering several performance indicators when making resource-related decisions [21]. However, computation offloading differs from joint wireless resource allocation involving power, bandwidth, user association, throughput, latency, fairness, and energy efficiency. These studies collectively establish the suitability of multi-objective optimisation for complex engineering resource-management problems. However, multi-objective optimisation alone cannot continuously learn adaptive allocation policies from changing wireless environments. MOOP is therefore incorporated into the proposed framework as the decision-criterion layer, while DRL provides adaptive decision-making and the predictive and stability components account for anticipated environmental variation and temporal consistency.

Reactive resource allocation primarily relies on the current network state and may consequently respond only after congestion, channel degradation, or QoS deterioration has occurred. Predictive resource management offers an alternative by incorporating anticipated network conditions into the allocation process. Predictive optimal channel allocation has demonstrated how forecasting future network conditions can improve resource decisions before those conditions materialise [6]. However, the corresponding framework was developed specifically for integrated sensing, communication, and computation environments and did not combine predictive allocation with DRL-based learning and stability-aware control within a heterogeneous cellular network. Edge-AI resource-recommendation techniques for 5G/6G channels have also highlighted the growing importance of AI-enabled resource recommendations in future wireless environments [26]. Although this work provides a broader perspective on intelligent resource management, it does not formulate the complete decision process as an integrated DRL–prediction–stability framework.

The transition from reactive communication towards proactive environmental awareness has also been emphasised through wireless environmental information theory for online and proactive environmental intelligence communication in 6G [36]. Nevertheless, that work does not provide the specific resource-allocation architecture developed in the present study. Fairness-aware deep Q-learning has additionally been applied to resource allocation and reconfigurable intelligent surface (RIS) optimisation for high-altitude platform networks, directly demonstrating the relevance of fairness to autonomous resource management [8]. However, the HAP/RIS environment differs from the heterogeneous cellular setting examined here, while predictive environmental modelling and stability-aware optimisation are not explicitly integrated. The literature therefore reinforces the importance of predictive and proactive intelligence while leaving scope for embedding prediction directly within a multi-objective DRL-based resource-allocation decision loop.

The growing heterogeneity of 5G and 6G services has made QoS-aware resource management increasingly important. QoS-aware network slicing and resource management have been examined for the Internet of Vehicles in 5G networks, demonstrating the importance of incorporating service requirements directly into resource-management decisions [12]. However, that application primarily concerns network slicing and vehicular communications rather than general heterogeneous wireless resource allocation. Intelligent congestion control with adaptive resource management has also been studied for 5G URLLC software-defined networks [31]. Although relevant because resource decisions respond to congestion and QoS conditions, the approach focuses predominantly on congestion control rather than simultaneously balancing throughput, latency, fairness, and energy efficiency.

QoS-aware resource sharing and power control have further been investigated for underlay many-to-many vehicular communications, demonstrating the importance of jointly managing resources and transmission power to improve QoS [16]. Nevertheless, the vehicular communication scenario differs from the multi-tier heterogeneous network considered in the present study. Broader examinations of deep learning for 6G have identified opportunities and challenges associated with intelligent wireless

systems [17], while performance evaluation and optimisation research has considered 6G from the perspective of key performance indicators (KPIs), analytical tools, and AI models [15]. These contributions reinforce the broader transition towards AI-driven and performance-aware 6G networks but do not provide the integrated autonomous resource-allocation mechanism introduced here. Collectively, the literature confirms the necessity of incorporating QoS requirements into intelligent network management. However, a unified mechanism capable of learning allocation policies, predicting future conditions, balancing multiple QoS criteria, and regulating temporal instability remains insufficiently developed.

Energy efficiency and fairness constitute two particularly important resource-management objectives that may conflict with throughput maximisation. The application of DRL to energy-efficient wireless communications has demonstrated that learning-based methods can improve communication performance while accounting for energy consumption [5]. Energy savings have likewise been considered alongside coverage optimisation through a multi-objective evolutionary framework [4]. The significance of simultaneously evaluating multiple performance indicators in resource-management decisions has also been demonstrated through multi-index evaluation learning for computation offloading [21]. In addition, fairness has been explicitly incorporated into deep Q-learning for resource allocation and RIS optimisation [8]. These findings establish the importance of energy efficiency and fairness as resource-management objectives. Nevertheless, existing approaches frequently address these criteria independently or within specialised scenarios rather than explicitly managing the trade-offs among throughput, latency, fairness, and energy efficiency in dynamic heterogeneous wireless networks. The proposed framework therefore integrates these criteria within a unified multi-objective decision formulation.

Adaptive learning introduces an additional challenge in dynamic wireless environments because resource-allocation decisions may vary considerably between successive decision intervals in response to changing channel conditions, traffic demand, user mobility, and learning behaviour. Maximising instantaneous performance alone is therefore insufficient; temporal consistency and stable network operation must also be maintained throughout the decision process. Research on resource allocation in software-defined networks has identified important challenges associated with intelligent resource management [32]. Machine-learning-based resource allocation and dynamic placement for virtualised network functions have further demonstrated the importance of adaptive resource decisions under changing network conditions [10]. The importance of robustness in intelligent optimisation for dynamic engineering systems has additionally been demonstrated through a hierarchical deep-learning and distributionally robust framework, as shown in Table 1 [14].

The remainder of this paper is organised as follows. Section 2 critically reviews existing research on DRL-based resource allocation, multi-objective optimisation, predictive resource management, stability-aware optimisation, and intelligent decision-support approaches. Section 3 describes the heterogeneous wireless system model and formulates resource allocation from a multi-criteria decision-making perspective. Section 4 introduces the proposed autonomous decision-support framework, encompassing the DRL decision engine, PEM, MOOP, and SAO mechanisms. Section 5 details the simulation environment, experimental configuration, baseline methods, and evaluation methodology. Section 6 presents and discusses the simulation results. Section 7 examines the managerial and practical implications of the proposed framework. Section 8 outlines the study limitations and future research directions. Finally, Section.

Table 1
Critical Compassion of Existing Resource Allocation Approaches and Identification of the Research Gap

Seq.	Ref.	Study Focus	Method/ Approach	Main Objective	Main Strength	Main Limitation Relative to this Study	Relevance to the Proposed Study
1.	Alwarafy et al. [1]	HetNet resource allocation	DRL survey	Review autonomous RRA	Provides a comprehensive overview of DRL-based resource allocation and management in next-generation heterogeneous networks	Survey-based study; does not propose or experimentally validate an integrated autonomous allocation framework	Establishes the importance of DRL for autonomous resource allocation
2.	Diamanti et al. [5]	Wireless communication	DRL	Energy-efficient communication	Demonstrates the effectiveness of DRL for energy-aware wireless resource optimization	Focuses on energy-efficient communication within a specific wireless scenario; limited consideration of multiple QoS criteria	Supports the use of DRL and energy efficiency as decision criteria
3.	Dutta et al. [6]	ISCC systems	Predictive optimization	Channel allocation	Introduces predictive information to support proactive channel allocation decisions	Does not integrate DRL-based adaptive learning with stability-aware optimization	Motivates the predictive component of the proposed framework
4.	She et al. [29]	6G edge computing	Multi-agent DRL	Task offloading	Provides adaptive and distributed learning for dynamic resource-related decisions	Primarily addresses computational task offloading rather than heterogeneous wireless resource allocation; limited joint QoS optimization	Supports adaptive and distributed decision-making in 6G environments
5.	Naguib et al. [23]	6G SDN	DRL	Data transmission optimization	Demonstrates AI-based adaptive optimization for data transmission in 6G software-defined networks	Limited integration of multiple conflicting objectives and predictive environmental information	Supports DRL-based autonomous decision-making in 6G networks
6.	Haider et al. [12]	5G IoV	QoS-aware network slicing	Resource management	Explicitly incorporates QoS requirements into resource management	Application-specific to vehicular networks and network slicing; does not jointly optimize fairness,	Supports QoS-aware resource allocation

						energy efficiency, and throughput	
7.	Sittakul et al. [32]	5G URLLC SDN	Reinforced Dueling DQN	Congestion/resource management	Addresses congestion and QoS requirements in latency-sensitive 5G networks	Primarily focuses on congestion control and URLLC; lacks predictive and stability-aware multi-objective integration	Supports latency- and QoS-aware intelligent resource management
8.	Benatia et al. [4]	Edge WiFi	Multi-objective evolutionary optimization	Energy and coverage optimization	Explicitly addresses trade-offs between energy consumption and network coverage	Does not incorporate adaptive DRL or future-state prediction; focuses on deployment optimization	Supports the use of multi-objective optimization and energy efficiency
9.	Kha and Quyet [19]	IRS-assisted wireless	MOOP	Energy/sensing/multicast optimization	Provides a multi-objective formulation for balancing several performance requirements	Designed for a specific IRS-assisted multicast and ISAC scenario; lacks autonomous DRL-based decision learning	Supports the MOOP formulation of conflicting wireless objectives
10	Ejaz et al. [7]	Cloud-edge-IoT	Multi-objective scheduling	Application scheduling	Considers multiple competing objectives in complex edge-cloud resource scheduling	Focuses on application scheduling rather than wireless resource allocation and does not model wireless-channel dynamics	Demonstrates the applicability of multi-objective decision-making in complex networked systems
11.	Etheredge et al. [8]	HAP/RIS	FAIR-DQL	Fairness-aware resource allocation	Explicitly incorporates fairness into DRL-based resource allocation and RIS optimization	Does not integrate future-state prediction and temporal stability into the allocation decision process	Supports fairness-aware DRL resource allocation
12.	Noor et al. [25]	5G/6G channels	Edge-AI recommendation	Resource recommendation	Highlights AI-enabled resource recommendation for 5G/6G environments	Does not establish an integrated closed-loop framework combining prediction, learning, multi-objective optimization, and stability	Supports AI-enabled resource intelligence
13.	Zhang	6G	Proactive	Environmental	Promotes proactive	Does not provide a	Motivates the

	[37]		environmental intelligence	awareness	environmental intelligence for future 6G communication	complete resource allocation mechanism or integrate environmental prediction with DRL and stability control	predictive/proactive component
14.	Proposed study	B5G/6G HetNet	MOOP + DRL + PEM + SAO	Throughput + latency + fairness + energy efficiency	Integrates adaptive learning, future-state prediction, multi-objective decision-making, and stability-aware control within one framework	Requires higher computational complexity and comprehensive validation for practical deployment	Provides an integrated intelligent decision-support framework for autonomous multi-criteria resource allocation

Although these studies demonstrate the importance of adaptive and robust intelligent decision mechanisms, SAO in the present study serves as an additional decision layer specifically intended to regulate variations between consecutive wireless resource-allocation actions. Stability is therefore not positioned as an alternative to DRL; instead, it complements adaptive learning by limiting abrupt changes in allocation decisions while preserving multi-objective performance requirements [15]. The critical assessment of the literature reveals four major research gaps. First, despite the strong adaptive decision-making capabilities demonstrated by existing DRL-based approaches, many remain application-specific, addressing problems such as task offloading, congestion control, vehicular communication, or energy-efficient wireless transmission.

Second, although multi-objective approaches effectively represent trade-offs among competing objectives, their integration with continuous DRL-based policy learning and future-state prediction remains limited. Third, predictive resource management facilitates proactive decision-making, yet prediction is not consistently incorporated into a closed-loop wireless resource-allocation mechanism that simultaneously accounts for decision stability. Fourth, existing QoS-aware and fairness-aware approaches address selected performance dimensions, but the joint consideration of throughput, latency, fairness, and energy efficiency under dynamic network conditions remains insufficient [16].

These gaps motivate an integrated autonomous decision-support framework in which DRL enables adaptive policy learning, PEM provides future-state awareness, MOOP captures conflicting decision criteria, and SAO regulates the temporal stability of resource-allocation decisions. Accordingly, wireless resource allocation is formulated as a continuous multi-criteria decision-making problem rather than an isolated optimisation task. The proposed study distinguishes itself from the reviewed approaches by combining four complementary capabilities within a single decision loop. DRL learns adaptive policies through interactions with the network environment, PEM supplies proactive information concerning anticipated network conditions, MOOP structures the competing decision criteria, and SAO controls the temporal evolution of allocation decisions [17]. The resulting framework therefore functions as an intelligent decision-support mechanism for autonomous resource management in heterogeneous B5G/6G networks.

The comparative analysis presented in Table 1 demonstrates that previous studies have addressed several important but largely individual dimensions of intelligent wireless resource management. DRL-based research has established the effectiveness of adaptive learning for resource allocation, transmission optimisation, and fairness-aware decision-making [1; 5; 8; 22; 28; 31]. Other studies have employed multi-objective optimisation to reconcile competing requirements involving energy consumption, coverage, sensing, and communication performance Benatia et al. [4]; [7; 18]. Predictive and proactive approaches have further demonstrated the value of incorporating anticipated network conditions into resource-management decisions [6; 26; 36]. Nevertheless, the reviewed literature generally concentrates on one or two of these capabilities rather than combining them within a unified autonomous decision-making process. In particular, the joint integration of adaptive DRL-based decision-making, predictive environmental information, multi-objective QoS optimisation, and temporal stability remains insufficiently explored. Moreover, numerous existing techniques are designed for specific application contexts, including vehicular networks, RIS-assisted systems, high-altitude platform (HAP) networks, edge-cloud environments, and task-offloading scenarios, thereby limiting their direct applicability to general heterogeneous cellular networks. The proposed study addresses these limitations by integrating DRL, PEM, MOOP, and SAO within a unified intelligent decision-support framework for heterogeneous B5G/6G wireless networks. Resource allocation is consequently treated as a continuous multi-criteria decision-making problem that jointly considers throughput, latency, fairness, and energy efficiency subject to network and QoS constraints.

2. DESCRIPTION SYSTEM

The proposed system model represents a realistic next-generation HetNet operating under highly

dynamic and uncertain conditions [17]. It provides the underlying environment within which the autonomous resource-allocation problem is formulated and addressed. The network adopts a hierarchical multi-tier architecture comprising macro cells that provide wide-area coverage and small cells that enhance capacity in densely populated areas. Such multi-layer deployments are fundamental to 5G and subsequent wireless generations because they simultaneously support coverage expansion and capacity densification. Within this environment, numerous mobile users compete for shared physical resources while generating heterogeneous traffic demands according to their respective applications.

Users are categorised into three service classes: eMBB, characterised by high data-rate requirements; URLLC, requiring extremely low latency and high reliability; and mMTC, involving very large numbers of low-power devices transmitting relatively small volumes of data. To reflect realistic network behaviour, the system is modelled as a dynamic and non-stationary environment whose operating conditions evolve over time. These variations arise from user mobility, wireless-channel fading and shadowing, and interference generated by simultaneous transmissions across overlapping cells. Consequently, resource-allocation decisions must be made under continuously changing and uncertain conditions rather than within a predictable and static network environment.

The system explicitly defines the principal communication resources as controllable decision variables, including radio spectrum in the form of frequency bands or sub-channels, transmission-power levels assigned to individual users or base stations, and time-slot assignments for data transmission. Because these resources are finite, their distribution across users and cells must be carefully coordinated to maximise network utilisation [18]. The central modelling challenge is therefore to formulate resource allocation as a real-time decision-making problem in which the network state is continuously observed and resource assignments are adaptively adjusted. The resulting allocation process seeks to optimise overall network performance while satisfying heterogeneous QoS requirements, maintaining fairness among users, and ensuring efficient utilisation of limited wireless resources within a highly dynamic and interference-rich environment.

2.1 System Overview and Modelling Perspective

The considered HetNet is formulated as a multi-tier wireless system comprising a dense population of users with heterogeneous service requirements served collectively by macro base stations and small cells. The operating environment evolves dynamically as user distributions, traffic demands, channel conditions, and interference levels vary over time. To represent the system mathematically, the following elements are introduced:

- A set of users UA
- A set of base stations UB
- A resource allocation vector including power, bandwidth, and user association decisions

The system exhibits inherent cross-layer coupling, whereby physical-layer variables, including SINR and transmission power, directly influence network-layer performance metrics such as throughput, latency, fairness, and energy efficiency.

2.1.1 User Association Model (Network Layer Coupling)

Each user is required to establish a connection with exactly one base station, represented mathematically through a binary association variable.

$$x_{u,b} \in \{0,1\} \quad uA \in U, \quad b \in B$$

$$x_{u,b} = 1, \text{if user uuu is associated with base station b}$$

$$x_{u,b} = 0, \text{otherwise}$$

The Uniqueness Constraint Ensures Stable Network Connectivity

$$\sum_{b\in B} x_{u,b} = 1 \qquad AU \in U \tag{1}$$

This constraint introduces a combinatorial structure, rendering the problem NP-hard because user association must be jointly optimised with power and resource allocation [19].

2.1.2 Physical Layer Model Signal-to-Interference-Plus-Noise Ratio (SINR)

SINR: This metric characterises the quality of a received wireless signal by expressing the useful signal power relative to the combined interference and noise power. It is mathematically expressed as:

$$\gamma_u = \frac{P_{u,b} h_{u,b}}{\sum_{j\neq b} P_{u,j} h_{u,j} + \sigma^2} \tag{2}$$

- γ_u SINR of user u.
- $P_{u,b}$ transmit power from serving base station b.
- $h_{u,b}$ channel gain between BS b and user u.
- $P_{u,j}$ transmit power from interfering base station j.
- $h_{u,j}$ interference channel gain.
- σ^2 dditive white Gaussian noise (AWGN) power

MOOP provides a mathematical framework for simultaneously optimising multiple performance objectives that are frequently conflicting [20; 21]. Unlike classical optimisation, which generally considers a single objective, such as maximising throughput or minimising cost, MOOP accommodates several interdependent objectives within the same formulation. This characteristic makes it particularly suitable for wireless communication systems, where network performance depends on multiple interacting criteria. For instance, resource allocation may seek to maximise throughput and energy efficiency while simultaneously minimising latency and maintaining fairness among users. These objectives inherently involve trade-offs because improving one performance dimension can adversely affect another. Increasing transmission power, for example, may enhance throughput while reducing energy efficiency and increasing interference, whereas imposing stringent fairness requirements may decrease overall system capacity.

Consequently, a single solution that simultaneously achieves the best value for every objective generally does not exist. Instead, MOOP identifies a collection of optimal trade-off solutions known as the Pareto-optimal set. A solution is Pareto optimal when no objective can be improved without degrading at least one other objective. Various techniques can be employed to solve MOOP, including weighted-sum methods, ε-constraint methods, evolutionary algorithms such as NSGA-II, and advanced machine-learning approaches such as DRL [22-24]. These methods seek to approximate the Pareto front, which represents the best attainable trade-offs among competing objectives. MOOP therefore provides a comprehensive framework for modelling and solving complex engineering problems in modern wireless networks, particularly when several performance criteria must be jointly optimised subject to practical system constraints. Mathematically, MOOP is expressed as:

$$maxF(x) = (F1(x) + F2(x) \ldots . FK(x) \tag{3}$$

x is the decision vector which includes the variables related to resource allocation, including transmission power, bandwidth assignment and user association, with each Fk(x) representing an individual objective function. The overall objective for the proposed system can be expressed in terms of a weighted sum representation:

$$maxJ(x) = \sum_{k=1}^{K} w_K \, F_K(x) \tag{4}$$

Where wK denotes the importance weight assigned to each objective. The optimization is performed under practical system constraints:

$$\mathrm{Pi} \leq \mathrm{P_i^{max}} \quad \mathrm{Ai} \ , \qquad \sum_{\mathrm{u}\in\mathrm{U}} \mathrm{Bu} \leq \mathrm{Btotal} \ , \mathrm{QoS_u} \geq \mathrm{QoS_u^{min}} \quad \mathrm{Au}$$

These constraints ensure compliance with power budgets, available bandwidth, and QoS requirements. Given the conflicting nature of the objectives, however, no single solution can simultaneously optimise all performance metrics. The system therefore seeks Pareto-optimal solutions, whereby improving one objective necessarily results in the deterioration of at least one other objective [25; 26].

2.2 Deep Reinforcement Learning (DRL) Layer

The DRL layer constitutes the core intelligent decision-making component of the proposed framework, with each base station operating as an autonomous decision unit capable of adapting to continuously evolving network conditions. Rather than relying on predefined rules or fixed optimisation models, each DRL agent progressively learns an effective decision policy through repeated interactions with the wireless environment [28]. At each time step, the agent observes the prevailing network state, which encompasses information such as traffic load, user distribution, interference levels, and channel quality. Based on these observations, it selects a resource-allocation action involving decisions such as user scheduling, bandwidth assignment, or transmission-power adjustment. The environment subsequently transitions to a new state and generates a reward reflecting the quality of the selected action.

The learning process seeks to maximise cumulative reward over time rather than merely optimising immediate performance. This enables the agent to develop long-term allocation strategies that jointly account for throughput, latency, energy efficiency, and fairness. Accordingly, the reward function integrates these objectives within a unified formulation, encouraging the agent to identify an appropriate overall trade-off rather than optimise an individual performance metric in isolation. Because the relationship between network states and desirable allocation actions is highly complex and nonlinear, deep neural networks are employed to approximate the underlying policy or value function. Their capacity to process high-dimensional inputs enables the agent to generalise its learned behaviour across different network conditions.

A key capability of the DRL layer is its adaptability to dynamic and non-stationary wireless environments. Continuous variations in user mobility, channel fading, and interference patterns require allocation policies to evolve as network conditions change. This adaptation is achieved through iterative learning, whereby accumulated experience informs subsequent decisions and progressively refines the policy. During this process, the agent balances exploration and exploitation by occasionally testing alternative actions to identify potentially superior strategies while also applying knowledge acquired from previous interactions. Within a multi-cell HetNet, individual DRL agents are typically associated with separate base stations and operate concurrently. Although direct interaction among agents may be limited, their decisions remain implicitly coupled because an action selected by one agent modifies the network environment subsequently observed by others. Consequently, each agent encounters an effectively non-stationary learning environment, increasing the complexity of policy optimisation. Despite this challenge, DRL remains suitable for the proposed framework because policies can be continuously updated in response to observed environmental changes.

Overall, the DRL layer formulates resource allocation as a sequential decision-making process in which previous interactions are exploited to improve future network performance. It therefore provides the adaptability and flexibility required for modern wireless systems without requiring every aspect of network dynamics to be explicitly represented through fixed mathematical optimisation models [29; 30]. As the intelligent decision-making engine of the framework, the DRL layer enables each base station to operate as an autonomous entity that progressively learns effective resource-

allocation behaviour through continuous interaction with the wireless environment. At time step t, the agent observes the current network state as:

$$St = \{SINR, Qt, Ut, It, Dt\} \quad (5)$$

Where SINR is signal-to-interference-plus-noise ratio, Qt is queue state, Ut is user distribution, It is interference conditions, and Dt is traffic demand. Based on this state, the agent selects an action:

$$\alpha t = \{Pt, Bt, At\} \quad (6)$$

Where Pt, Bt , and At denote power allocation, bandwidth assignment, and scheduling decisions. The objective is to maximize the expected cumulative discounted reward:

$$\mathrm{Rt} = \sum_{\mathrm{r=t}}^{\infty} \gamma^{\mathrm{r-t}} \mathrm{r_t} \quad (7)$$

Where γ is the discount factor and rt is the instantaneous reward. The reward is designed as a multi-objective utility:

$$\mathrm{rt} = \alpha \mathrm{Tt} + \mathrm{BLt} + \delta \mathrm{Et} + \eta \mathrm{Ft} \quad (8)$$

Where Tt is throughput, Lt is latency, Et is energy efficiency, and Ft is fairness, $\alpha, B, \delta, and \ \eta$ balancing coefficients. The policy is approximated using deep neural networks:

$$\pi\theta\ (\alpha t, St)$$

Where $\boldsymbol{\theta}$ represents the trainable network parameters. Through this learning process, the agent progressively adapts to the dynamic and nonlinear wireless environment without requiring an explicit analytical model.

2.3 Predictive Environment Model

PEM constitutes an important component of the autonomous resource-allocation framework by enabling decisions to incorporate anticipated network states rather than relying exclusively on current observations [31]. HetNet environments are inherently dynamic and non-stationary, with user locations, channel quality, traffic demand, and interference patterns continuously evolving over time. Decisions derived solely from the current network state may therefore become short-sighted, particularly under high user mobility and bursty traffic conditions. Incorporating predictive capability allows resource allocations to be adjusted proactively, thereby reducing the likelihood of performance degradation before adverse network conditions materialise.

PEM learns temporal patterns from historical and real-time network information. Relevant observations may include user mobility trajectories, SINR variations, traffic-arrival rates, and handover events. Using time-series forecasting or lightweight supervised-learning techniques, the predictive module estimates network conditions over a short future horizon. These estimates may encompass anticipated channel conditions for individual users, prospective congestion within particular cells, and expected interference fluctuations arising from neighbouring transmissions. Incorporating such future-state information into the decision process enables the resource-allocation agent to evaluate both present and anticipated network conditions, supporting more stable and forward-looking actions. This capability is particularly important for services with stringent latency or reliability requirements, where delayed reactions to deteriorating conditions may result in QoS violations.

PEM can also improve the stability of learning-based mechanisms, particularly reinforcement-learning processes operating within non-stationary environments. Abrupt changes in state distributions can destabilise learning and slow convergence. Short-term prediction can reduce the uncertainty associated with these variations and provide more consistent information to the learning agent. Consequently, policy updates can become more reliable, while learned policies may generalise more effectively across changing network conditions. Computational efficiency is equally important for real-

time implementation. PEM therefore employs simplified or lightweight predictive techniques rather than computationally intensive architectures, balancing forecasting accuracy against processing latency. This consideration is particularly important in large-scale deployments where multiple base stations must operate concurrently under limited computational resources [32; 33].

Conceptually, PEM is a computational representation that learns, estimates, or simulates how an environment evolves over time according to its internal dynamics, external conditions, and actions. Its principal purpose is to anticipate future system states and enable subsequent decisions to incorporate those predictions. Within engineering and intelligent systems, PEM therefore functions as a virtual representation of the operating environment, allowing the system to move beyond purely reactive responses to current observations and instead anticipate likely future conditions. Such predictive capability is particularly valuable in autonomous systems, wireless networks, robotics, and resource management. Conceptually, PEM can be expressed as:

$$\mathrm{S_{t+1}} = \mathrm{f(St, At, Et)} \tag{9}$$

Where St current state of the environment, At action taken by the agent/ system, Et is external contextual factors , S_{t+1} is predicted next state, and $f(\,..)$ is learned or designed predictive.

2.4 Stability-Aware Optimization

The primary purpose of SAO is to regulate the evolution of resource-allocation decisions so that they remain stable, smooth, and reliable within time-varying wireless environments, particularly under highly dynamic conditions characterised by rapid network changes [9]. In such settings, decisions generated by a learning-based controller may fluctuate substantially between consecutive time steps because of exploration, stochastic channel variations, or abrupt changes in user demands. Excessive variations can adversely affect system performance by increasing signalling overhead, disrupting the consistency of user experience, and potentially causing QoS violations. SAO therefore introduces a controlled mechanism that limits undesirable temporal fluctuations in allocation decisions while preserving satisfactory network performance.

The underlying principle is to formulate resource allocation as a temporally coupled decision problem rather than as a sequence of independent optimisation problems. Accordingly, a decision at the current time step is determined not only by prevailing network conditions but also with consideration of the allocation selected at the preceding time step. Mathematically, this temporal consistency is achieved by imposing constraints on variations in the decision vector over time. Assume at time t the decision vector is {x_t} = (xt1, xt2, ..., xtK) of power allocation, bandwidth assignment and user association, respectively, then the stability would be enforced by limiting the difference between two consecutive decisions. This constraint ensures that resource-allocation decisions evolve smoothly over time, preventing abrupt transitions or discontinuities between consecutive configurations.

Rather than permitting substantial changes at each decision interval, the system restricts adjustments to a reasonable range relative to the preceding configuration. Stability is incorporated into the optimisation process through both regularisation and explicit constraints. The original objective, which typically seeks to maximise network utility in terms of throughput, energy efficiency, and fairness, is augmented with a penalty term that discourages large deviations from previous decisions. The resulting formulation therefore balances performance improvement against temporal stability, allowing substantial allocation changes only when they provide sufficient gains to justify the associated disruption.

A more rigorous characterisation of stability can be established through Lyapunov theory, which provides a formal framework for analysing the temporal behaviour of dynamic systems. A Lyapunov function quantifies deviation from a stable operating condition, while its temporal variation is represented by the Lyapunov drift. By controlling this drift while simultaneously optimising the relevant performance objectives, the system can be directed towards stable operation. This produces a drift-

plus-penalty formulation that jointly promotes system stability and network efficiency. Consequently, excessive oscillations or divergence can be mitigated even when the network operates under time-varying conditions.

Queue dynamics constitute another important dimension of stability, particularly for delay-sensitive services. Incoming data packets must be served sufficiently rapidly to prevent unbounded queue growth. This behaviour is represented through queue-evolution equations that describe changes in queue length according to packet arrivals and service rates. Within this context, stability requires the expected queue length to remain finite over time, thereby contributing directly to bounded delay. Incorporating queue-stability constraints consequently enables the system to maintain QoS requirements, particularly those associated with low-latency communications.

SAO additionally accounts for uncertainty arising from factors such as channel-estimation errors and unexpected interference variations. Rather than optimising exclusively for a single deterministic network condition, the formulation can be extended to accommodate worst-case conditions within predefined uncertainty bounds. SAO therefore promotes smooth, reliable, and temporally consistent resource-allocation decisions. Because abrupt variations in control actions within learning-based systems may increase signalling overhead and degrade user experience [27; 34], temporal regularisation is incorporated into the optimisation process. Accordingly, the following decision-smoothness constraint is imposed:

$$\|xt - x_{t-1}\| \leq \in$$

where ϵ controls the maximum allowable variation. A regularized objective is then formulated:

$$Jstable\ (xt) = J(xt) - \lambda\|xt - x_{t-1}\|^2 \quad (10)$$

where λ determines the strength of stability enforcement. To provide theoretical guarantees, a Lyapunov-based formulation is introduced. The Lyapunov function is defined as:

$$\mathrm{L(Qt)} = \frac{1}{2}\,\mathrm{Q}^2 \quad (11)$$

The corresponding drift is:

$$\Delta(Qt) = E[L(Qt+1) - L(Q)|Qt] \quad (12)$$

The optimization minimizes the drift-plus-penalty term:

$$\min \Delta(\mathrm{Qt}) - \mathrm{V}\ \mathrm{E}[\mathrm{rt}|\mathrm{Qt}] \quad (13)$$

where V controls the trade-off between performance and stability. Queue evolution is modeled as:

$$\mathrm{Qt} + 1 = \max(\mathrm{Qt} - \mathrm{ut}, 0) + \mathrm{At} \quad (14)$$

Where At packet arrivals, and ut is service rate. Bounded queue length provides latency stability and QoS.

This optimisation approach strengthens system performance across operating conditions that may differ from those assumed during optimisation, thereby improving transferability and reducing sensitivity to modelling inaccuracies. SAO consequently refines the decisions generated by the learning component by transforming resource allocation into a controlled dynamic process that jointly considers performance optimisation, temporal consistency, and robustness. Achieving this balance between high system efficiency and reliable, stable long-term operation is essential for practical deployment [35; 36].

From a decision-support perspective, the autonomous resource-allocation problem considered in this study is formulated as a multi-criteria sequential decision-making problem within a dynamic and non-stationary heterogeneous wireless environment. Rather than addressing resource allocation as an isolated optimisation task, the proposed framework establishes a continuous decision process in which the network state is observed, relevant information is evaluated, and allocation actions are selected

according to multiple, potentially conflicting performance requirements. The principal decision variables comprise transmission power, bandwidth allocation, user association, and scheduling, whereas the decision criteria encompass throughput, latency, fairness, energy efficiency, and QoS. These criteria are evaluated jointly because improving one performance dimension may adversely affect another. For example, increasing transmission power can enhance received signal quality and throughput while simultaneously increasing interference and reducing energy efficiency. Similarly, imposing more stringent fairness requirements may restrict the aggregate throughput achievable by the network. Resource allocation therefore requires explicit mechanisms for evaluating and balancing competing objectives rather than optimising a single performance measure independently.

The base stations are connected to the DRL decision-making unit, which functions as the adaptive decision-maker within the framework. At each decision interval, the agent observes the network state, including SINR, queue status, user distribution, interference conditions, and traffic demand, and subsequently selects an appropriate resource-allocation action. MOOP provides the multi-criteria decision structure by representing the trade-offs among the considered network objectives, whereas DRL enables allocation policies to be learned through continuous interaction with the wireless environment. The decision process is further supported by PEM, which uses historical and current network information, including user mobility trajectories, SINR variations, traffic-arrival rates, and handover events, to anticipate the evolution of network conditions. These predictions are combined with observations of the current network state, enabling the allocation mechanism to account for expected environmental changes rather than responding solely to instantaneous conditions. Such predictive capability is particularly important within the considered non-stationary environment, where user mobility, channel variations, traffic fluctuations, and interference may substantially affect the effectiveness of previously selected allocation decisions.

Temporal Consistency (TC) is incorporated into the decision process through SAO. Each selected resource-allocation action is evaluated relative to the preceding allocation state, while temporal regularisation and stability constraints restrict excessive variations between consecutive decisions. Queue evolution is also incorporated to support stable network operation and compliance with QoS requirements. Consequently, the proposed decision process does not pursue instantaneous network utility alone; it simultaneously accounts for the temporal behaviour and stability of successive resource-allocation decisions [37; 38].

3. SIMULATION ENVIRONMENT

The simulation parameters were selected to establish a realistic yet computationally efficient 5G HetNet environment for evaluating the proposed MOOP framework. A simulation area of 1000 × 1000 m^2 was adopted to represent a typical urban deployment in which macro base stations, small cells, and mobile users coexist under overlapping coverage conditions. The macro base stations provide large-scale coverage, while the twelve small cells establish an appropriate balance between network coverage and capacity while generating meaningful inter-cell interference and load distribution. A total of 200 users was considered to represent moderate-to-high network traffic, creating sufficient competition for radio resources to evaluate user association, resource allocation, and fairness under dynamic operating conditions.

A carrier frequency of 3.5 GHz was selected because of its widespread use in 5G New Radio (NR) deployments, while a system bandwidth of 100 MHz provides a practical balance among coverage, spectral efficiency, and expected throughput. Propagation conditions were represented using the 3GPP Urban Macro (UMa) path-loss model, Rayleigh fading, and 8 dB log-normal shadowing, thereby accounting for large-scale attenuation, multipath propagation, and environmental obstructions characteristic of urban wireless environments. User mobility was represented through the Random Waypoint model, with velocities ranging from 3 to 30 km/h to capture pedestrian and slow-moving vehicular scenarios. Consequently, channel quality, interference conditions, and handover events

evolve continuously throughout the simulation.

Transmission powers of 46 dBm and 30 dBm were assigned to macro base stations and small cells, respectively, consistent with practical 5G deployment settings and providing an appropriate balance among coverage extension, interference management, and energy consumption, as illustrated in Figure 2 and summarised in Table 2. The thermal-noise spectral density was set to the standard room-temperature noise floor to support accurate SINR estimation. Three standardised 5G service classes were incorporated: eMBB, URLLC, and mMTC. Their heterogeneous traffic characteristics and conflicting QoS requirements enable resource allocation to be evaluated as a multi-objective optimisation problem.

Proximal Policy Optimization (PPO) was adopted because of its stable convergence behaviour and suitability for dynamic environments. The policy and value functions were represented using three-layer deep neural networks containing 256 neurons per layer, providing sufficient modelling capacity to capture complex nonlinear relationships among network states, channel conditions, traffic demands, and resource-allocation decisions. The learning rate was set to 1×10^{-4}, the batch size to 256, and the discount factor to 0.99, while training was conducted over 1000 episodes to achieve stable policy convergence within a manageable computational duration. A 1 ms time interval, corresponding to the Transmission Time Interval (TTI) employed in practical 5G systems, enables the learning agent to continuously observe changing network conditions and adapt resource-allocation decisions accordingly. The complete simulation environment and experimental configuration are illustrated in Figure 2, while the parameters employed throughout the experiments are summarised in Table 2.

Table 2
The Parameters used throughout the Simulations are Summarized

Seq.	Parameter	Value
1.	Simulation Area	$1000 * 1000\ m^2$
2.	Number of Macro BSs	3
3.	Number of Small Cells	12
4.	Number of Users	200
5.	Carrier Frequency	$3.5\ GHz$
6.	System Bandwidth	$100\ MHz$
7.	Channel Model	Rayleigh Fading
8.	Path Loss Model	3GPP Urban Macro
9.	Shadowing	$8\ dB$
10.	Mobility Model	Random Waypoint
11.	User Speed	$3 - 30\ Km/h$
12.	Maximum BS Power	$46\ dBm$
13.	Small Cell Power	$30\ dBm$
14.	Noise Spectral Density	$174\ dBm/Hz$
15.	QoS Classes	eMBB, URLLC, mMTC
16.	DRL Algorithm	PPO
17.	Neural Network	3 Hidden Layers
18.	Hidden Neurons	256
19.	Learning Rate	$1 * 10^{-4}$
20.	Batch Size	256
21.	Replay Buffer	100000
22.	Discount Factor γ	0.99
23.	Number of Episodes	1000
24.	Time Slot	$1\ mS$

The simulation environment was configured to provide a realistic yet computationally tractable representation of a dynamic 5G HetNet for evaluating the proposed MOOP framework. The network

consists of macro base stations, small cells, and mobile users operating under overlapping coverage conditions. A 1000 × 1000 m² simulation area containing 3 macro base stations and 12 small cells was considered. This configuration provides sufficient spatial diversity and inter-cell interference to assess user association, resource allocation, QoS provisioning, and fairness under heterogeneous network conditions. A population of 200 mobile users was employed to generate moderate-to-high traffic demand and create adequate competition for the available radio resources.

The carrier frequency and system bandwidth were set to 3.5 GHz and 100 MHz, respectively, representing a typical 5G New Radio (NR) operating configuration. Wireless propagation was represented using the 3GPP Urban Macro (UMa) path-loss model together with Rayleigh small-scale fading and 8 dB log-normal shadowing. Collectively, these models capture large-scale propagation losses, multipath effects, and signal variations caused by obstacles within urban environments, as illustrated in Figure 2. User movement was represented through the Random Waypoint model, with velocities ranging from 3 to 30 km/h. This range captures pedestrian and low-speed vehicular mobility while introducing continuous changes in channel conditions, interference levels, user associations, and handover requirements. Maximum transmission powers were fixed at 46 dBm for macro base stations and 30 dBm for small cells, allowing the framework to evaluate trade-offs among coverage, interference, throughput, and energy consumption. A thermal noise spectral density of (-174) dBm/Hz was incorporated into the SINR calculation.

The simulation incorporates three representative 5G service classes: eMBB, URLLC, and mMTC. These classes were selected because their heterogeneous and potentially conflicting QoS requirements provide an appropriate setting for multi-objective resource allocation. eMBB users primarily require high throughput, whereas URLLC users demand low latency and reliable resource access, and mMTC involves a large number of connections with comparatively low data-rate requirements. Their simultaneous presence therefore enables resource allocation to be formulated as a multi-objective problem involving throughput, latency, fairness, and energy efficiency. Proximal Policy Optimization (PPO) was employed as the DRL algorithm within the proposed autonomous resource-allocation framework.

Its clipped policy-update mechanism was selected to promote stable policy optimisation by restricting excessively large policy changes during training, which is particularly relevant to the considered HetNet where network states and resource demands evolve continuously. Fully connected deep neural networks (DNNs), comprising three hidden layers with 256 neurons per layer, were used to implement the PPO policy and value functions. This architecture provides sufficient representational capacity to learn the nonlinear relationships among network states, channel conditions, traffic requirements, QoS constraints, and resource-allocation decisions without introducing unnecessary computational complexity. The learning rate was set to 1×10^{-4}, the mini-batch size to 256, and the discount factor (gamma) to 0.99. Training was conducted for 1000 episodes.

At every 1-ms Transmission Time Interval (TTI), the DRL agent observes a snapshot of the current network state containing information on channel conditions, traffic, user association, and resource utilisation. Resource-allocation actions are then selected according to this observed state. Following each action, the environment evaluates network performance and returns a reward corresponding to the multi-objective requirements. The reward formulation jointly considers throughput, latency, fairness, and energy efficiency, thereby encouraging allocation decisions that improve overall network performance rather than maximising an individual metric. This multi-objective formulation is particularly important within the considered HetNet because improvements in throughput may increase latency and/or energy consumption, whereas aggressive reductions in energy consumption may adversely affect throughput and/or QoS satisfaction. PEM anticipates changes in relevant network conditions, while SAO regulates the resulting resource-allocation decisions to limit excessive temporal fluctuations. The interaction among PPO, PEM, and SAO therefore enables adaptive resource allocation under time-varying network conditions.

To facilitate experimental reproducibility, the network-level simulation settings and principal PPO training parameters are explicitly reported in Table 1. These include the network dimensions, number of base stations, user population, propagation models, mobility assumptions, transmission powers, QoS classes, neural-network architecture, learning rate, batch size, discount factor, number of training episodes, and TTI. Because PPO is an on-policy reinforcement-learning algorithm, the proposed implementation does not employ a conventional experience-replay buffer of the type commonly used by off-policy methods. Instead, on-policy trajectories are collected during training and subsequently used for the corresponding PPO policy and value-function updates.

Training performance was assessed by examining the evolution and stabilisation of accumulated episode rewards across the 1000 training episodes. Following training, the final learned policy was used for the performance evaluation presented in the Results and Discussion section. All performance metrics were obtained using the same network configuration specified in Table 2, thereby maintaining consistency between the training and evaluation environments. Where available, software versions, hardware configuration, random seeds, and total computational time should be reported separately because these factors can influence the exact reproducibility of stochastic DRL experiments. These implementation-specific details are therefore distinguished from the network and learning parameters reported in Table 2.

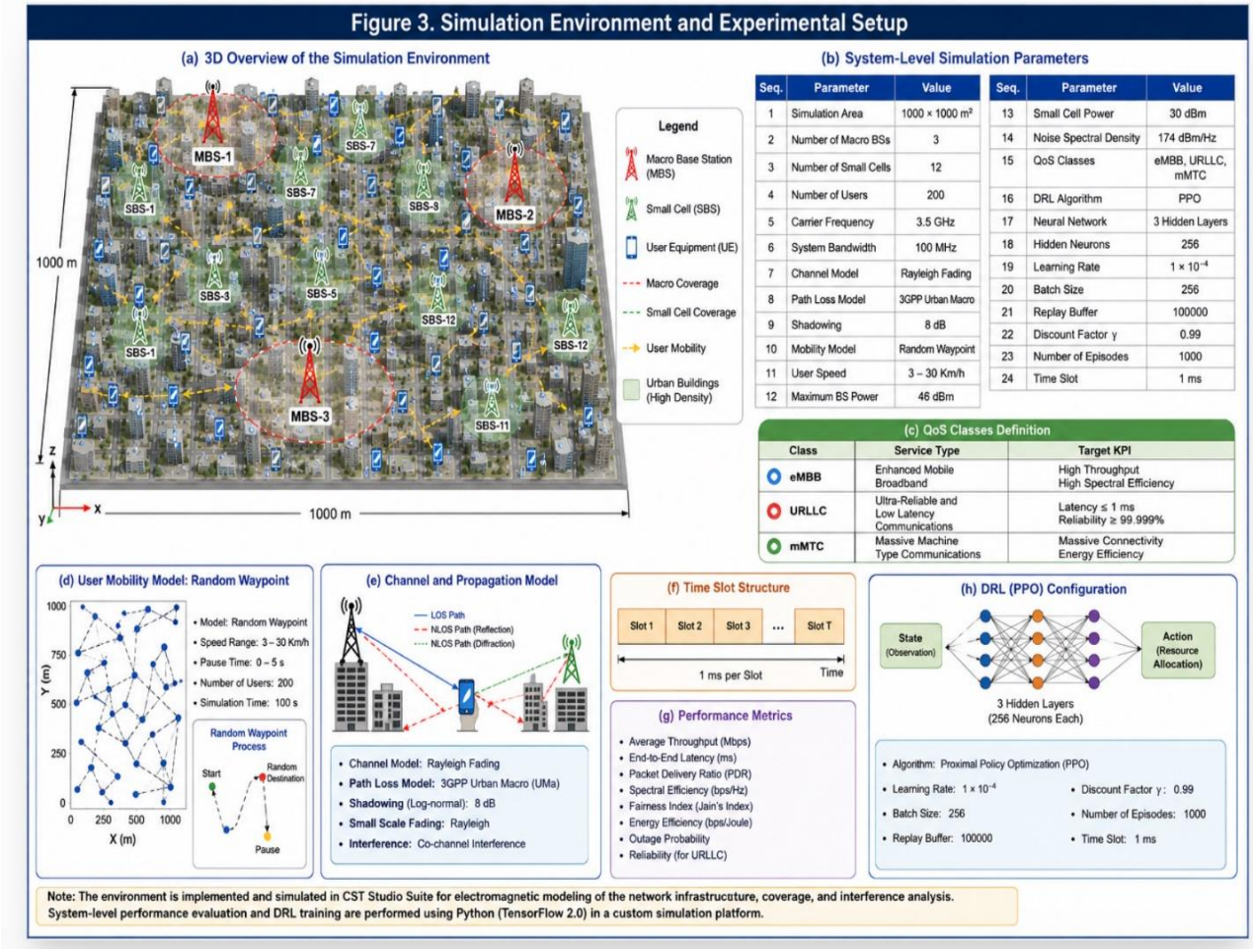


Fig.2: System-Level Simulation Environment and Experimental Configuration for the Proposed DRL-Based 5G HetNet Framework

4. RESULTS AND DISCUSSION

The simulation results demonstrate the effectiveness of the proposed integrated framework for

autonomous resource allocation in heterogeneous wireless networks. As illustrated in Figure 3, MOOP, PEM, DRL, and SAO are incorporated within a unified decision-making process. Rather than relying on an individual optimisation strategy, the framework enables these components to operate collaboratively, allowing resource-allocation decisions to account for both current network conditions and anticipated environmental changes. This integration provides a balanced and scalable approach to resource management in dynamic 5G/B5G wireless environments.

The simulation findings further demonstrate the complementary contribution of each framework component. MOOP jointly represents throughput, latency, fairness, and energy efficiency as competing optimisation objectives. Unlike conventional single-objective approaches that may improve one performance metric at the expense of others, the proposed formulation seeks an operating point capable of satisfying multiple QoS requirements simultaneously. The optimisation process therefore moves towards Pareto-efficient solutions that reconcile conflicting performance objectives while maintaining stable network operation. PEM contributes to this process by providing future-state awareness within the dynamic network environment.

Wireless conditions are inherently non-stationary because user mobility, channel fading, traffic fluctuations, and interference levels vary continuously. Resource-allocation approaches based exclusively on instantaneous network observations may consequently respond only after performance deterioration has occurred. In contrast, PEM anticipates forthcoming network states, allowing the controller to identify potential traffic congestion, channel degradation, and interference variations before they significantly affect network performance. Resource allocation can therefore be adjusted proactively, reducing the likelihood of QoS deterioration under increasing traffic loads or rapid mobility conditions.

The adaptive capability of DRL is also demonstrated through the simulation process. The cumulative reward exhibits a stable convergence pattern over successive training episodes, indicating progressive improvement in the learned resource-allocation policy through repeated interaction with the wireless environment. Unlike rule-based optimisation approaches that require explicit modelling of network conditions, DRL learns the relationships among observed network states, allocation actions, and long-term rewards. Consequently, the framework can adapt its allocation strategy across variations in channel quality, user density, traffic demand, and interference while progressively improving its decision policy throughout training. SAO provides an additional mechanism for controlling temporal variations in resource-allocation decisions. Learning-based resource-management systems may produce abrupt changes in bandwidth allocation, transmission power, or scheduling decisions during exploration, substantial environmental transitions, or training. Such oscillatory behaviour can increase signalling overhead and reduce decision consistency.

By incorporating temporal regularisation and Lyapunov-based stability control, SAO limits excessive variations between successive scheduling intervals and promotes smoother allocation decisions. This mechanism supports stable queue dynamics, mitigates congestion under heavy traffic conditions, and improves overall network stability without compromising the ability of the framework to respond to changing network conditions. The interaction among these four components contributes collectively to improved network performance. Adaptive spectrum utilisation combined with interference-aware resource allocation enables more efficient use of available radio resources and consequently improves bandwidth efficiency.

Proactive congestion management and stable queue control reduce end-to-end latency, thereby supporting delay-sensitive services. At the same time, the multi-objective formulation promotes more equitable resource distribution among users while limiting unnecessary transmission-power consumption whenever network conditions permit, contributing to improved energy efficiency. The simulations further demonstrate the ability of the proposed framework to operate under dynamically changing conditions. Integrating prediction, adaptive learning, and stability control enables the system to respond to variations in user mobility, traffic demand, and interference. PEM continuously

anticipates forthcoming conditions, DRL progressively adapts the allocation policy through interaction with the environment, and SAO regulates successive decisions to limit undesirable temporal fluctuations. Their coordinated operation consequently supports more reliable resource allocation during long-term network operation. The closed-loop framework also provides advantages over conventional resource-allocation approaches under the operating conditions defined by the parameters in Figure 2.

Static allocation methods offer limited flexibility when wireless conditions change rapidly, whereas reactive heuristic approaches may respond only after resource-allocation performance has deteriorated. In contrast, the proposed framework combines prediction, adaptive decision-making, multi-objective optimisation, and stability control within a unified process, enabling more consistent network performance across the simulated operating conditions. Overall, the simulation findings validate the integration of MOOP, PEM, DRL, and SAO as a scalable and intelligent framework for autonomous resource management in future heterogeneous wireless networks. By jointly incorporating multi-objective decision-making, future-state awareness, adaptive policy learning, and temporal stability, the proposed approach supports diverse QoS requirements while enabling efficient and reliable resource allocation for next-generation communication services.

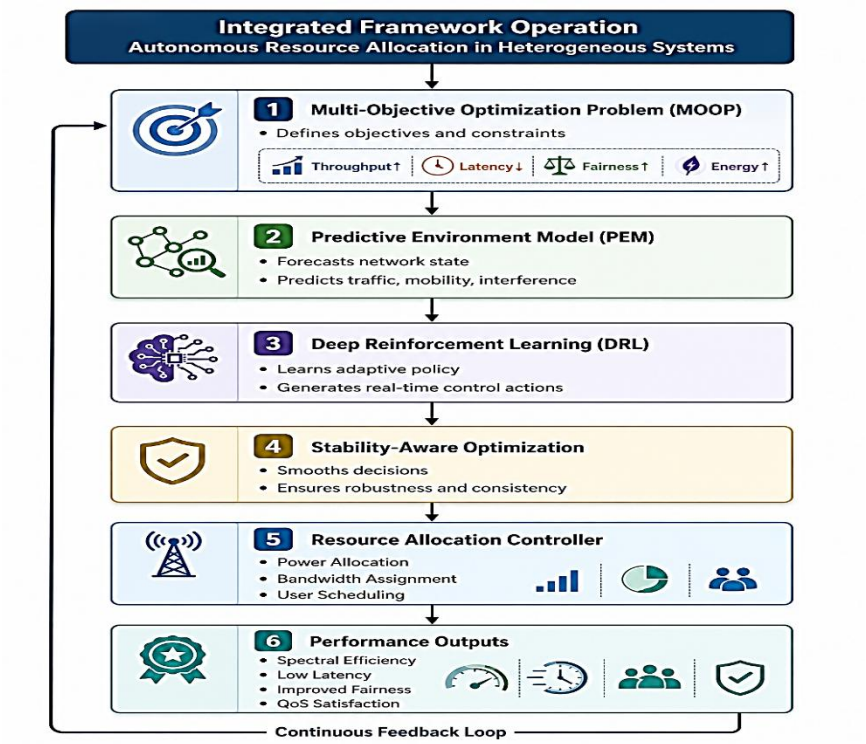


Fig.3: Integrated Multi-Layer Intelligent Framework for Autonomous Resource Allocation with Predictive Modelling and Deep Reinforcement Learning in Heterogeneous Networks (MOOP)

Figure 4 compares the predictive performance of LSTM, GRU, Transformer, and the proposed PEM using Mean Absolute Error (MAE) across increasing time steps. All four models exhibit progressively decreasing MAE as the number of time steps increases, indicating that prediction accuracy improves as additional temporal observations become available during training. Nevertheless, substantial differences are evident in their convergence behaviour and final prediction errors. Among the evaluated models, LSTM exhibits the highest initial prediction error and the slowest convergence. Its MAE remains comparatively high throughout the observation period, indicating a relatively limited ability to capture the rapidly changing temporal dynamics of the considered wireless environment. GRU consistently achieves lower MAE than LSTM across almost all time steps, demonstrating improved

predictive performance. Its gated architecture enables temporal dependencies to be learned effectively while retaining a comparatively simple model structure. Transformer further reduces prediction error relative to both recurrent models. Its attention-based architecture facilitates the modelling of temporal relationships across longer observation windows, resulting in faster convergence and lower MAE.

The proposed PEM, however, consistently achieves the lowest prediction error throughout the evaluated time steps. Its MAE decreases rapidly during the initial training period before gradually converging towards a lower and more stable error level. This behaviour indicates both improved prediction accuracy and stable convergence compared with the other evaluated models. At the end of the 1000 time steps, the approximate MAE values for LSTM, GRU, and Transformer are 0.18, 0.14, and 0.10, respectively, whereas PEM achieves an MAE of approximately 0.05, as illustrated in Figure 4. Relative to LSTM, this corresponds to an approximate 72% reduction in prediction error, calculated as (0.18-0.05)/0.18 x 100 ≈72%. The quantitative comparison therefore demonstrates a substantial improvement in predictive accuracy over the recurrent and attention-based models evaluated in this study.

The consistently lower MAE achieved by PEM is particularly relevant to the subsequent resource-allocation process. More accurate forecasts of future network conditions enable the controller to anticipate variations in traffic demand, channel conditions, and interference before they significantly affect network performance. PEM therefore extends the decision process beyond instantaneous network observations and supports proactive resource-allocation decisions. Its smoother convergence behaviour further indicates relatively stable predictive performance as additional temporal observations become available, which is desirable for integration with the DRL-based decision-making process. Overall, Figure 4 demonstrates that PEM provides the strongest predictive performance among the evaluated models, exhibiting the lowest absolute prediction error, faster convergence, and smoother learning behaviour. The approximately 72% reduction in MAE relative to LSTM quantitatively confirms the improvement in prediction accuracy. These findings support the integration of PEM as the predictive component of the proposed autonomous resource-allocation framework, where improved future-state estimation can facilitate more reliable, proactive, and stable allocation decisions under dynamic HetNet conditions.

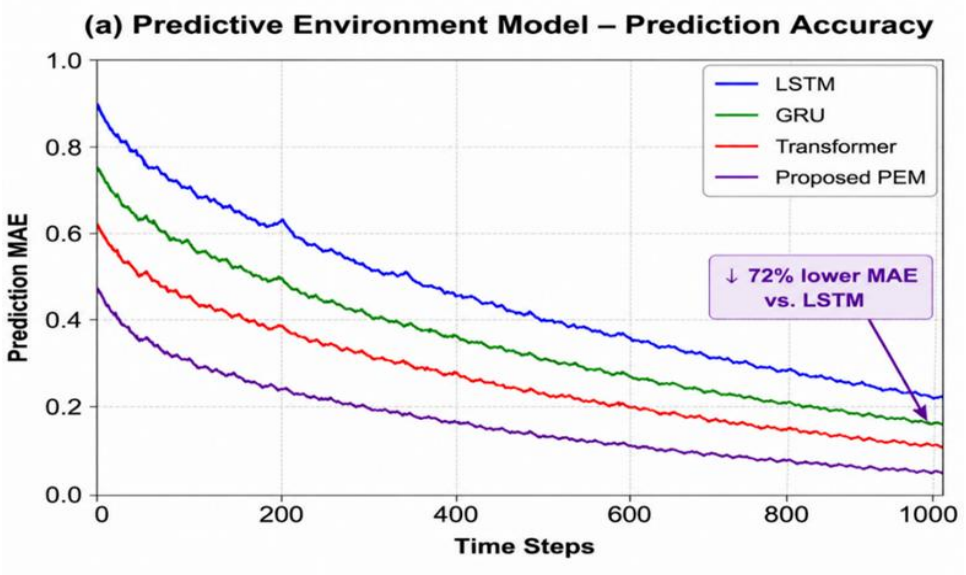


Fig.4: Prediction Performance of the Proposed Predictive Environment Model Compared with Baseline Deep Learning Methods

Figure 5 illustrates the cumulative rewards achieved by the evaluated DRL algorithms across 1,000

training episodes. Cumulative reward serves as an indicator of the quality of the resource-allocation policies learned by the agents, with higher values representing a more favourable long-term balance among the optimisation objectives incorporated into the reward function. The negative rewards recorded during the initial episodes correspond to the exploration stage, when the agents possess limited knowledge of the network environment and consequently select comparatively inefficient resource-allocation actions. As training progresses, cumulative rewards increase across all evaluated algorithms, demonstrating progressive adaptation to changing network conditions and refinement of the learned decision policies.

Despite this general improvement, the algorithms exhibit markedly different convergence characteristics. DQN produces the lowest cumulative reward and the slowest improvement, indicating comparatively limited learning efficiency for the considered dynamic HetNet resource-allocation problem. DDPG achieves higher rewards than DQN and demonstrates stronger adaptation to the network environment. PPO further improves learning performance, exhibiting faster cumulative-reward growth and attaining a higher final reward than both DQN and DDPG. The proposed DRL framework achieves the highest cumulative reward throughout most of the training process and converges comparatively rapidly towards a stable, high-reward policy. This behaviour indicates that the proposed decision-making mechanism learns resource-allocation policies capable of achieving a more favourable overall trade-off among the considered QoS objectives. The higher cumulative reward reflects an improved ability to jointly account for network conditions, resource availability, traffic requirements, and QoS priorities rather than optimising an individual performance criterion in isolation.

The performance difference between the proposed framework and the conventional DRL algorithms becomes increasingly apparent as training progresses. After approximately 600–800 episodes, the reward curves begin to level off, indicating that the learning processes are approaching convergence and that additional training produces comparatively smaller policy improvements. Nevertheless, the proposed framework maintains the highest cumulative reward during the later training episodes, demonstrating superior learned-policy performance under the considered simulation conditions.

Overall, the results support the effectiveness of the proposed framework as an autonomous decision-making mechanism for wireless resource allocation. By combining multiple decision criteria with adaptive DRL-based learning, the agent progressively identifies allocation actions that improve long-term network utility. However, the observed differences should be interpreted as simulation-based performance comparisons rather than evidence of statistical significance. Where applicable, results from multiple independent training runs, together with appropriate dispersion measures and statistical significance analyses, are reported in the final manuscript to strengthen the reliability of the comparison.

Figure 5: Comparison of DRL Algorithms Based on Cumulative Reward Performance Over Training Episodes.

Figure 6 illustrates the convergence behaviour and stability of the considered optimisation methods as the number of iterations increases. The baseline method exhibits the slowest convergence and achieves the lowest final objective-function value, indicating comparatively limited optimisation efficiency relative to the learning-based approaches. Conventional DRL without stability awareness improves upon the baseline by converging more rapidly and attaining a higher objective value. In contrast, the proposed DRL approach incorporating stability awareness achieves both the fastest convergence and the highest objective value among the evaluated methods.

The improved performance of the proposed approach is attributable to the explicit incorporation

of stability awareness into the optimisation process, which facilitates the identification of high-performance solutions while limiting undesirable oscillations between successive decisions. The proposed method also exhibits a smoother convergence trajectory, indicating greater consistency across iterations and smaller fluctuations in the objective-function value. Its rapid improvement during the initial iterations, followed by a gradual progression towards a stable converged value, further demonstrates its ability to efficiently identify favourable solutions while maintaining system stability. Overall, the results presented in Figure 6 demonstrate that integrating stability awareness within the DRL framework improves convergence efficiency while simultaneously enhancing the stability and reliability of the optimisation process.

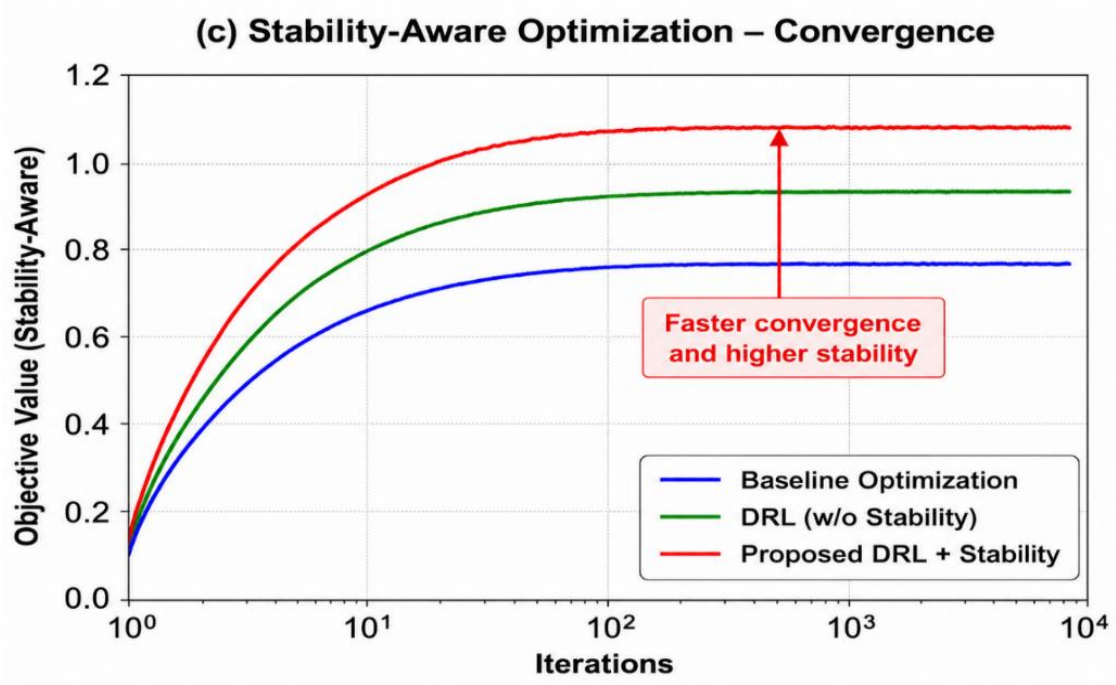


Fig.6: Convergence of Optimization Methods with Stability Considerations

Figure 7 illustrates the evolution of average SINR across the considered time steps for the different optimisation methods. The baseline heuristic approach exhibits the lowest SINR throughout most of the evaluated horizon, with only gradual improvement and a final value of approximately 8 dB. This behaviour indicates its limited ability to adapt to changing network conditions and to effectively coordinate interference management and communication-resource allocation. DRL without PEM provides a substantial improvement over the baseline, with average SINR increasing more rapidly during the initial time steps before gradually stabilising at approximately 12–13 dB. This improvement demonstrates the advantage of learning-based resource optimisation over the conventional heuristic approach.

Incorporating PEM into DRL further improves communication-link quality. The DRL + PEM scheme increases the average SINR to approximately 17–18 dB, as shown in Figure 7, indicating more effective resource management and interference mitigation. The proposed DRL + PEM + Stability framework achieves the highest average SINR, reaching approximately 23–24 dB by the end of the evaluated time horizon. It also exhibits rapid improvement during the initial time steps followed by comparatively stable behaviour thereafter. At the final time step, the proposed framework provides an SINR improvement of approximately 9.3 dB relative to the baseline heuristic method.

The superior SINR performance of the proposed framework reflects the combined contribution of DRL-based optimisation, predictive information from PEM, and stability-aware decision control. In particular, stability awareness promotes a more consistent performance trajectory by limiting excessive

variations in successive resource-allocation decisions. Although minor fluctuations remain because of the dynamic nature of the network environment, the proposed approach maintains the highest SINR throughout the evaluated period. Overall, Figure 7 demonstrates that incorporating stability awareness into the DRL + PEM framework improves communication-link quality while supporting more stable optimisation behaviour, resulting in superior SINR performance compared with the baseline heuristic and DRL-based benchmark schemes.

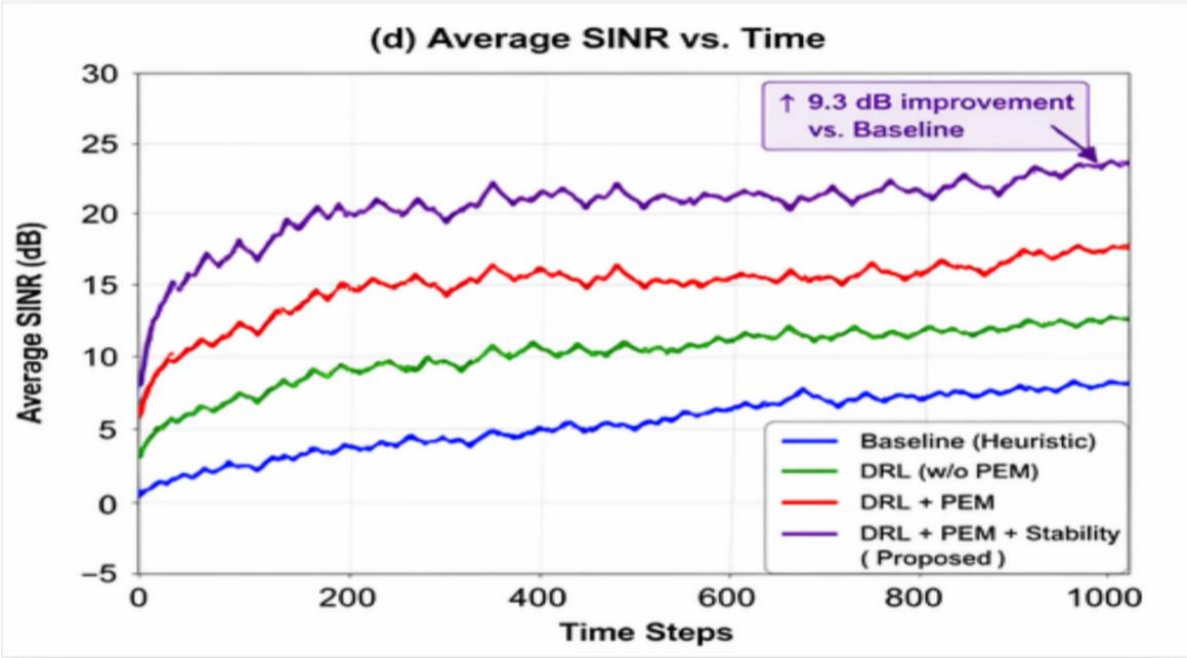


Fig.7: Average SINR Improvement Over Time with Stability-Aware DRL

Figure 8 illustrates the evolution of the average data rate across the considered time steps for the evaluated optimisation approaches. The average data rate reflects the effectiveness of each method in utilising available communication resources to support data transmission under dynamic network conditions. The baseline heuristic method exhibits the lowest performance throughout the evaluated period, increasing gradually from approximately 11 Mbps to approximately 50 Mbps. This relatively modest improvement indicates limited gains in resource utilisation over time.

DRL without PEM achieves substantially higher data rates than the baseline, increasing from approximately 24 Mbps to approximately 75 Mbps. This improvement demonstrates the advantage of learning-based optimisation in adapting resource-allocation decisions to changing network conditions. Incorporating PEM into DRL further enhances performance, with the DRL + PEM approach reaching approximately 103 Mbps by the end of the evaluated period, as shown in Figure 8. Its consistently higher data rate relative to both the baseline heuristic and DRL-only approaches indicates more effective coordination and utilisation of communication resources. The proposed DRL + PEM + Stability framework achieves the highest average data rate among all evaluated approaches. Its performance increases rapidly during the initial time steps, followed by sustained improvement along a comparatively stable trajectory, ultimately reaching approximately 130–132 Mbps. The proposed framework therefore maintains a clear performance advantage over the other methods throughout most of the considered time horizon.

Although minor fluctuations are evident because of the dynamic optimisation environment, the data-rate trajectories remain relatively smooth over time. Overall, Figure 8 demonstrates that incorporating PEM and stability awareness into the DRL framework substantially improves the achievable data rate compared with the baseline heuristic and DRL-only approaches. The results further

indicate that the proposed framework enhances data-transmission efficiency while maintaining a stable performance trajectory, supporting its suitability for resource allocation in dynamic wireless communication environments.

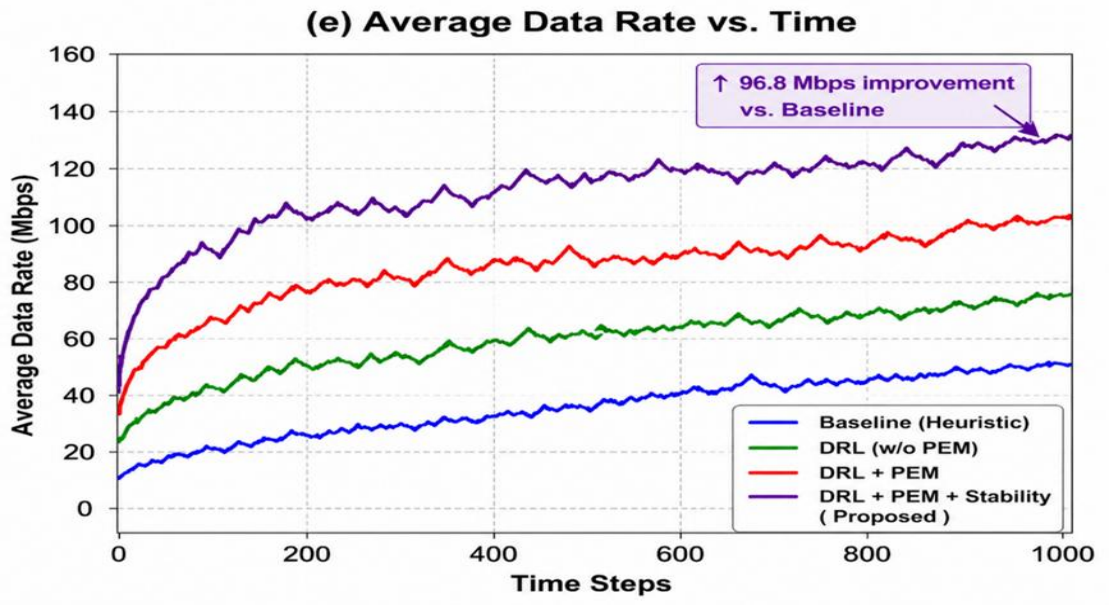


Fig.8: Average Data Rate Improvement Over Time with Stability-Aware DRL

Figure 9 illustrates the relationship between average data rate and average SINR for the four evaluated optimisation strategies. Each cluster represents performance samples associated with a particular method, while the fitted trend lines indicate the overall relationship between SINR and data rate. Across all four approaches, a positive association is evident, with higher SINR generally corresponding to increased data rates. The baseline heuristic method occupies the lowest-performing region of the plot, exhibiting comparatively low SINR and data-rate values. This distribution indicates relatively limited communication performance under the considered operating conditions. DRL without PEM shifts the performance samples towards higher SINR and data-rate levels, demonstrating the advantage of adaptive learning-based resource optimisation over the heuristic approach. Further improvement is observed with DRL + PEM, whose samples are concentrated at higher SINR and data-rate values than those of both the baseline and DRL-only methods. This shift indicates that incorporating PEM enables the framework to operate more frequently under favourable communication conditions.

The proposed DRL + PEM + Stability framework forms the highest-performing cluster in Figure 9, with its samples concentrated towards the upper-right region of the plot, corresponding to simultaneously higher SINR and data rates. The observations are also relatively concentrated around the fitted trend line, indicating a consistent relationship between these two performance measures across the evaluated operating conditions. The clear separation of the proposed approach from the remaining clusters further demonstrates its overall performance advantage. The positive relationship observed between SINR and data rate is consistent with the fundamental behaviour of wireless communication systems, whereby improved received signal quality generally supports higher achievable data rates. Under the idealised Shannon-capacity formulation, this relationship is represented by the logarithmic dependence of channel capacity on SINR. However, the empirical data-rate behaviour presented in Figure 9 is additionally influenced by resource-allocation decisions, interference conditions, system constraints, and the characteristics of the optimisation strategy. Accordingly, the results should be interpreted as empirical evidence of a positive SINR–data-rate relationship rather than as direct verification of the Shannon-capacity limit.

Overall, Figure 9 provides further evidence that the proposed stability-aware DRL framework operates within a more favourable performance region than the benchmark approaches, simultaneously achieving higher SINR and data rates. The comparatively compact distribution of its performance samples also indicates a more consistent performance pattern under the evaluated conditions, further supporting the effectiveness of integrating DRL, PEM, and stability awareness for reliable and efficient wireless resource optimisation.

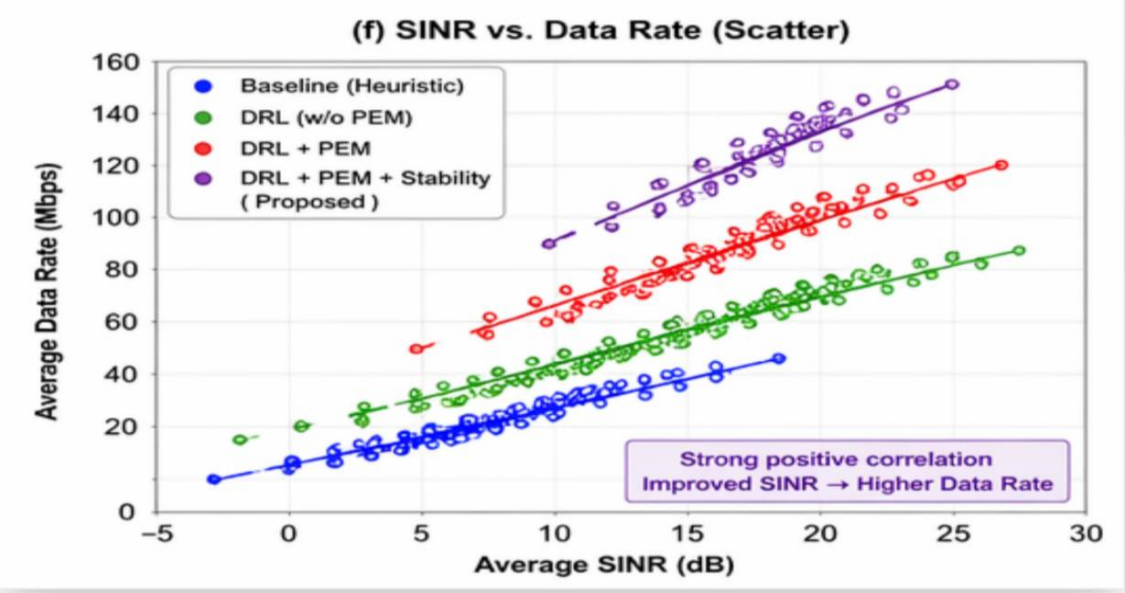


Fig.9: SINR–Data Rate Correlation: Performance Analysis of DRL-Based Methods

Figure 10 provides a comprehensive comparison of the four optimisation frameworks in terms of throughput, latency, fairness, and energy efficiency. Unlike the preceding individual performance analyses, this comparison summarises the overall behaviour of the evaluated approaches across complementary performance dimensions under the considered network conditions. The static framework exhibits the weakest overall performance, achieving a throughput of 78 Mbps, latency of 42 ms, fairness index of 0.71, and energy efficiency of 5.2 bits/J. These results indicate that fixed resource allocation has limited capability to adapt to changing network conditions and cannot simultaneously maintain high data-transmission performance, low latency, equitable resource distribution, and efficient energy utilisation.

The reactive framework improves upon the static approach, increasing throughput to 96 Mbps while reducing latency to 31 ms. Its fairness index rises to 0.79, and energy efficiency reaches 6.8 bits/J. These improvements demonstrate the benefit of adjusting resource-allocation decisions in response to changing network conditions rather than maintaining fixed allocations. Nevertheless, its performance remains constrained because reactive decision-making does not explicitly learn the long-term consequences of successive resource-allocation actions. DRL produces further improvements across all four metrics. Throughput increases to 118 Mbps, while latency decreases to approximately 23 ms. Simultaneously, the fairness index reaches 0.86 and energy efficiency increases to 8.1 bits/J. These results demonstrate that learning-based resource allocation can exploit observed network information to improve multiple performance dimensions concurrently, enabling a more favourable trade-off among transmission performance, delay, fairness, and energy utilisation.

The proposed framework achieves the strongest overall performance, with a throughput of 135 Mbps, latency of 16 ms, fairness index of 0.93, and energy efficiency of 9.7 bits/J. Relative to the static framework, these values correspond to an approximately 73% increase in throughput, 62% reduction in latency, 31% improvement in the fairness index, and 87% increase in energy efficiency. The proposed

framework also improves upon conventional DRL, increasing throughput from 118 Mbps to 135 Mbps and reducing latency from approximately 23 ms to 16 ms, while raising the fairness index from 0.86 to 0.93 and energy efficiency from 8.1 bits/J to 9.7 bits/J. These simultaneous improvements are particularly important because the considered performance criteria are inherently interdependent and partially conflicting. Increasing throughput may require greater resource utilisation and can intensify congestion, whereas improving fairness may necessitate redistributing resources among users. Similarly, reducing latency and increasing energy efficiency can impose additional constraints on resource scheduling. The results in Figure 10 therefore indicate that the proposed framework does not improve an individual metric at the expense of the remaining objectives; instead, it achieves a more balanced compromise across the four performance dimensions.

Overall, Figure 10 consolidates the findings of the preceding performance evaluations and demonstrates a consistent progression from the static and reactive approaches to DRL and, ultimately, the proposed framework. Among the evaluated methods, the proposed approach achieves the highest throughput, lowest latency, highest fairness index, and greatest energy efficiency. These quantitative results support the effectiveness of the proposed stability-aware framework in simultaneously improving communication performance, resource-utilisation efficiency, and allocation fairness, thereby demonstrating its suitability as a comprehensive multi-objective resource-allocation solution for dynamic wireless communication environments.

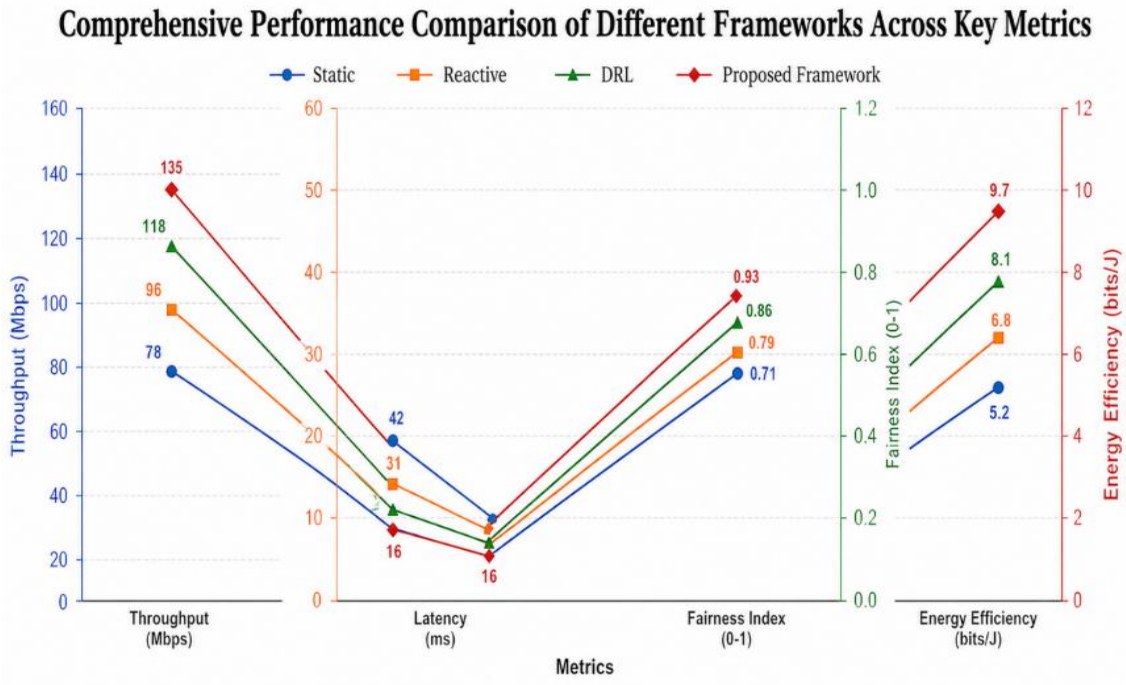


Fig.10: Comprehensive Multi-Metric Performance Comparison of Optimization Frameworks

The results presented in Figures 6–10 consistently demonstrate the performance advantages of the proposed stability-aware DRL framework over the evaluated benchmark approaches. The convergence analysis shows that incorporating stability awareness improves both the rate and consistency of optimisation, while the temporal evaluations indicate that the proposed framework sustains higher SINR and average data-rate levels throughout the considered time horizon. The SINR–data-rate analysis further demonstrates a positive relationship between communication quality and achievable transmission performance, with the proposed framework predominantly occupying the higher-performance region. The system-level comparison reinforces these findings by demonstrating simultaneous improvements in throughput, latency, fairness, and energy efficiency relative to the static, reactive, and conventional DRL approaches. Specifically, the proposed framework achieves a

throughput of 135 Mbps, latency of 16 ms, fairness index of 0.93, and energy efficiency of 9.7 bits/J, outperforming the benchmark methods across all evaluated performance dimensions. Importantly, these gains occur concurrently rather than through improvement in one metric at the expense of another, indicating more balanced resource-allocation behaviour.

Collectively, the findings demonstrate the complementary roles of DRL, PEM, and stability awareness within the proposed architecture. DRL provides adaptive decision-making capability, PEM contributes predictive information concerning anticipated network conditions, and stability awareness regulates the temporal evolution of resource-allocation decisions under dynamic operating conditions. The consistency observed across the convergence, SINR, data-rate, and multi-metric evaluations strengthens the evidence supporting the effectiveness of their integration. Under the considered simulation conditions, the proposed framework therefore provides more balanced and reliable optimisation performance than the evaluated benchmark methods when communication quality, transmission efficiency, latency, fairness, and energy utilisation are considered jointly.

To systematically assess the effectiveness of the proposed autonomous resource-allocation framework, four system-level strategies are compared: static allocation, reactive allocation, conventional DRL-based allocation, and the proposed DRL–PEM–SAO framework. These approaches represent progressively greater levels of decision-making capability, ranging from fixed allocation and state-responsive adaptation to learning-based resource management and, ultimately, predictive and stability-aware autonomous decision-making. A consistent performance progression is observed across these strategies. Static allocation produces the weakest overall results, achieving 78 Mbps throughput, 42 ms latency, a fairness index of 0.71, and energy efficiency of 5.2 bits/J. Its comparatively limited performance reflects the inability of a fixed allocation policy to respond to changing network conditions. Reactive allocation improves these values to 96 Mbps, 31 ms, 0.79, and 6.8 bits/J, respectively, demonstrating the benefit of adjusting decisions according to observed network-state changes. Nevertheless, its decisions remain primarily dependent on current conditions rather than the learned long-term consequences of resource-allocation actions. These findings are consistent with the results discussed for Figure 10.

Conventional DRL provides a further improvement, achieving 118 Mbps throughput, approximately 23 ms latency, a fairness index of 0.86, and energy efficiency of 8.1 bits/J. These results demonstrate the value of learning-based resource allocation, whereby information obtained from the network environment can be used to develop allocation policies that achieve a more favourable balance among throughput, latency, fairness, and energy efficiency. The proposed DRL–PEM–SAO framework extends these gains, attaining 135 Mbps throughput, 16 ms latency, a fairness index of 0.93, and energy efficiency of 9.7 bits/J. Its advantage therefore extends beyond an isolated performance dimension, with improvements occurring simultaneously across all four metrics. This characteristic is particularly relevant to HetNets because these objectives are inherently interdependent and may conflict with one another.

A direct comparison with conventional DRL more clearly demonstrates the additional contribution of the proposed framework. Throughput increases from 118 Mbps to 135 Mbps, corresponding to an improvement of approximately 14.4%, while latency decreases from approximately 23 ms to 16 ms, representing a reduction of approximately 30.4%. The fairness index increases from 0.86 to 0.93, equivalent to an improvement of approximately 8.1%, and energy efficiency rises from 8.1 bits/J to 9.7 bits/J, corresponding to an increase of approximately 19.8%. These results indicate that the observed performance gains extend beyond those attributable to DRL alone and support the contribution of the additional predictive and stability-aware mechanisms. Specifically, PEM supplies future-state information, whereas SAO promotes greater temporal consistency in the optimisation trajectory. Their integration with DRL consequently produces a more balanced resource-allocation process under the considered simulation conditions.

In addition to the system-level evaluation, the learning component was compared with three DRL

algorithms: DQN, DDPG, and PPO. The cumulative-reward analysis indicates that DQN exhibits the slowest learning progression, whereas DDPG achieves faster reward improvement but does not attain the performance level of the proposed DRL model. PPO demonstrates stronger learning behaviour and reaches a higher reward than the other conventional alternatives. The proposed DRL model achieves the highest cumulative reward and fastest convergence among the evaluated algorithms. The reported results indicate an approximately 28% higher reward relative to DDPG. Although the reward trajectories progressively approach a plateau as training advances, indicating convergence, the proposed DRL model maintains the highest performance level, as presented in Table 3.

Overall, the baseline comparisons demonstrate a systematic progression in autonomous decision-making capability. Static allocation relies on a fixed policy, reactive allocation introduces adaptation to observed network states, and conventional DRL incorporates learning-based decision-making. The proposed DRL–PEM–SAO framework advances this progression by integrating future-state information and temporal stability into the resource-allocation process. The resulting benefits are observed simultaneously across throughput, latency, fairness, and energy efficiency. The comparative findings therefore support the interpretation of the proposed architecture as an integrated autonomous decision-making framework rather than merely another DRL-based resource-allocation algorithm.

Table 3
Qualitative Comparison of DRL Algorithms Based on the Existing Simulation Results

DRL Algorithm	Learning Behaviour	Convergence	Cumulative Reward	Role in Evaluation
DQN	Slow learning progression	Slowest	Lowest among evaluated methods	Value-based baseline
DDPG	Faster reward growth	Moderate	Lower than Proposed DRL	Actor–critic baseline
PPO	Improved learning behaviour	Faster than DQN/DDPG	Higher than conventional alternatives	Policy-based baseline
Proposed DRL	Fastest learning	Fastest convergence	Highest; ≈28% above DDPG	Proposed decision engine

To quantify the contribution of the principal components of the proposed autonomous resource-allocation framework, a component-wise ablation analysis was conducted using the configurations already evaluated in the simulation study. The analysis progressively considers conventional DRL, DRL with PEM, and the complete framework obtained by additionally incorporating SAO. This incremental evaluation isolates the respective contributions of adaptive learning, predictive environmental awareness, and stability control to overall resource-allocation performance. Conventional DRL serves as the learning-only baseline without predictive environmental information. As illustrated in Figure 7, this configuration achieves an average SINR of approximately 12–13 dB towards the end of the considered time horizon. Incorporating PEM increases the average SINR to approximately 17–18 dB, demonstrating an additional performance gain from predictive environmental information beyond that achieved through DRL alone. The complete DRL–PEM–Stability configuration provides the highest average SINR, reaching approximately 23–24 dB. Its trajectory is also comparatively smoother, indicating that stability awareness contributes not only to the achieved performance level but also to greater temporal consistency in resource-allocation decisions.

A comparable incremental improvement is evident in the average data-rate results. The DRL-only configuration increases the average data rate from approximately 24 Mbps at the beginning of the evaluated interval to approximately 75 Mbps towards its end. With PEM incorporated, the final data rate increases to approximately 103 Mbps. The complete DRL–PEM–Stability framework achieves the highest value, reaching approximately 130–132 Mbps while maintaining a smoother temporal trajectory. These findings demonstrate that predictive information contributes to improved resource coordination and data-transmission efficiency, while the stability-aware mechanism further promotes

consistency in successive allocation decisions. The convergence results provide complementary evidence regarding the contribution of SAO. As shown in Figure 6, conventional DRL already improves convergence relative to the baseline approach; however, incorporating stability awareness produces faster convergence and a higher final objective value. The stability-aware configuration also follows a smoother convergence trajectory with smaller fluctuations, demonstrating greater consistency throughout the optimisation process.

The component-wise analysis therefore reveals a progressive contribution from each module. DRL establishes the fundamental adaptive decision-making capability, PEM introduces future-state information that improves SINR and data-rate performance, and SAO enhances convergence consistency while limiting undesirable temporal fluctuations. Consequently, the complete DRL–PEM–Stability configuration provides the strongest overall performance among the evaluated configurations. As demonstrated by the comprehensive multi-metric evaluation, the complete framework achieves a throughput of 135 Mbps, latency of 16 ms, fairness index of 0.93, and energy efficiency of 9.7 bits/J, as shown in Table 4. The contribution of stability awareness is particularly evident in the temporal behaviour of the decision process and cannot be characterised solely through SINR or endpoint throughput. The convergence analysis demonstrates that the stability-aware configuration reaches a higher objective value through a faster and smoother trajectory than DRL without stability awareness. The reduced fluctuations indicate that SAO regulates successive resource-allocation decisions and limits abrupt policy changes. Its principal contribution is therefore reflected in the temporal consistency and reliability of the decision-making process, complementing the performance improvements provided by DRL and PEM.

Table 4
Component-Wise Ablation Analysis of the Proposed Framework

Configuration	DRL	PEM	Stability Awareness	Final/Observed SINR	Final/Observed Data Rate	Main Observed Contribution
Baseline heuristic	✗	✗	✗	≈8 dB	≈50 Mbps	Reference performance
DRL-only	✓	✗	✗	≈12–13 dB	≈75 Mbps	Adaptive resource allocation
DRL + PEM	✓	✓	✗	≈17–18 dB	≈103 Mbps	Predictive resource management
Proposed DRL + PEM + Stability	✓	✓	✓	≈23–24 dB	≈130–132 Mbps	Adaptive + predictive + stable allocation

5. CONCLUSION

This study presented a stability-aware hybrid DSS for autonomous resource allocation in heterogeneous B5G/6G wireless networks. The framework integrates MOOP, DRL, PEM, and SAO to support adaptive, predictive, and temporally stable decision-making under dynamic network conditions. Unlike static or reactive approaches, it jointly considers throughput, latency, fairness, energy efficiency, SINR, and data rate while incorporating anticipated network states and decision stability. Simulation results demonstrate the effectiveness of the proposed framework, achieving 135 Mbps throughput, 16 ms latency, a fairness index of 0.93, and energy efficiency of 9.7 bits/J. Compared with static allocation, throughput and energy efficiency improve by approximately 73% and 87%, respectively, while latency decreases by approximately 62%, alongside improved fairness. Convergence, SINR, and data-rate results further demonstrate more consistent performance, while the progressive gains from conventional allocation to DRL, DRL with PEM, and the complete framework confirm the complementary contributions of adaptive learning, prediction, multi-objective

optimisation, and stability control.

From an engineering management perspective, the proposed DSS provides a coordinated multi-criteria mechanism for balancing communication performance, resource utilisation, fairness, energy efficiency, and operational consistency under changing QoS requirements. However, the findings remain specific to the adopted simulation conditions, and practical deployment requires further investigation of computational complexity, communication latency, prediction errors, scalability, and integration with existing network-management systems. Future research should evaluate the framework in ultra-dense, highly mobile, uncertain, and decentralised environments, including federated and distributed reinforcement learning. Extensions to network slicing, semantic communications, advanced predictive models, hardware-in-the-loop evaluation, and realistic 6G testbeds should also be considered. Explainable Artificial Intelligence (XAI) could further improve decision transparency, operator trust, and managerial oversight, supporting the development of a scalable, explainable, and practically deployable DSS for reliable and energy-efficient B5G/6G networks.

6. ACKNOWLEDGEMENT

The authors also acknowledge the University of Sumer as the bodies that supported the work by providing resources that supported the realization of this study. Colleagues in the Departments of Communication Engineering.

7. ETHICAL STATEMENT

This study does not contain any studies with human or animal subjects performed by any of the authors.

8. CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest to this work.

9. DATA AVAILABILITY STATEMENT

Data are available from the corresponding author upon reasonable request.

References

[1] Alwarafy, A., Abdallah, M., Ciftler, B. S., Al-Fuqaha, A., & Hamdi, M. (2021). Deep reinforcement learning for radio resource allocation and management in next generation heterogeneous wireless networks: A survey. *arXiv preprint arXiv:2106.00574*. https://doi.org/10.48550/arXiv.2106.00574

[2] Ariffin, A., Afifi, F., Zaki, F., Hanif, H., & Anuar, N. B. (2025). Federated learning in network traffic classification: Taxonomy of implementation, application, and impact on sixth-generation wireless networks. *Engineering Applications of Artificial Intelligence*, *158*, 111471. https://doi.org/10.1016/j.engappai.2025.111471

[3] Baek, Y., Yoon, G., Kim, S., Kang, B., Lee, S., Choi, M.-i., & Park, S. (2026). Prediction-guided deep reinforcement learning for energy-efficient HVAC control with PICVs in smart buildings. *Energy Reports*, *15*, 109040. https://doi.org/10.1016/j.egyr.2026.109040

[4] Benatia, M. A., Ben Abdelaziz, F., & Sahnoun, M. h. (2025). Energy savings and coverage optimization in edge WiFi sensor deployment for buildings: A multi-objective evolutionary approach. *Energy Economics*, *141*, 108096. https://doi.org/10.1016/j.eneco.2024.108096

[5] Diamanti, M., Kapsalis, G., Tsiropoulou, E. E., & Papavassiliou, S. (2023). Energy-Efficient Rate-Splitting Multiple Access: A Deep Reinforcement Learning-Based Framework. *IEEE Open Journal of the Communications Society*, *4*, 2397-2409.

https://doi.org/10.1109/OJCOMS.2023.3322047

[6] Dutta, A., Roy, A., Saxena, N., & Weber, S. (2024). Predictive Optimal Channel Allocation in Integrated Sensing, Communication and Computation Systems. 2024 International Symposium on Networks, Computers and Communications (ISNCC),

[7] Ejaz, M., Asim, M., Wani, M. A., & Shakil, K. A. (2026). FAIR-DQL: Fairness-Aware Deep Q-Learning for Enhanced Resource Allocation and RIS Optimization in High-Altitude Platform Networks. *Computers, Materials & Continua*, *86*(3), 29. https://doi.org/10.32604/cmc.2025.072464

[8] Etheredge, M., Aznar Poveda, J., Pedratscher, S., Younesi, A., & Fahringer, T. (2026). Pulse: Multi-objective scheduling of service-based applications in multi-cluster cloud–edge–IoT infrastructures. *Journal of Network and Computer Applications*, *248*, 104439. https://doi.org/10.1016/j.jnca.2026.104439

[9] Fotio Tiotsop, L., Servetti, A., Barkowsky, M., Pocta, P., Mizdos, T., Van Wallendael, G., & Masala, E. (2023). Predicting individual quality ratings of compressed images through deep CNNs-based artificial observers. *Signal Processing: Image Communication*, *112*, 116917. https://doi.org/10.1016/j.image.2022.116917

[10] Gadouh, K. A., & Sellami, H. K. (2024). Enhancing Virtualized Network Functions through Advanced Machine Learning for Efficient Resource Allocation and Dynamic Placement. 2024 IEEE/ACS 21st International Conference on Computer Systems and Applications (AICCSA),

[11] Gülmez, B. (2026). Artificial intelligence-driven smart grid optimization: A comprehensive review of machine learning, renewable integration, and cybersecurity applications. *Energy Conversion and Management: X*, *30*, 101849. https://doi.org/10.1016/j.ecmx.2026.101849

[12] Haider, A., Shahzadi, I., Kiran, S., & Shakoor, K. (2026). Chitosan Introduction. In A. Haider, I. Shahzadi, S. Kiran, & K. Shakoor (Eds.), *Chitosan Nanoparticles for Brucellosis Treatment* (pp. 1-16). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-16292-2_1

[13] Hamdi, W., Dağdeviren, O., & Bulut, H. (2025). QoS-aware Network Slicing and Resource Management for Internet of Vehicles in 5G networks. *Ad Hoc Networks*, *178*, 103976. https://doi.org/10.1016/j.adhoc.2025.103976

[14] Haroon, M. S., Abbas, Z. H., Muhammad, F., & Abbas, G. (2020). Analysis of coverage-oriented small base station deployment in heterogeneous cellular networks. *Physical Communication*, *38*, 100908. https://doi.org/10.1016/j.phycom.2019.100908

[15] Heydari, M., Niknam, T., & Parsa, N. (2026). A hierarchical deep learning and distributionally robust framework for urban multi-energy systems. *Results in Engineering*, *29*, 108533. https://doi.org/10.1016/j.rineng.2025.108533

[16] Islam, M. M., Hasan, K., & Jeong, S. H. (2026). Performance evaluation and optimization for 6G networks: A survey of KPIs, Tools, and AI models. *ICT Express*, *12*(2), 390-416. https://doi.org/10.1016/j.icte.2025.12.012

[17] Islam, T., & Noh, H. (2026). Access-rate enhancement for underlay many-to-many vehicular communications: QoS-aware resource sharing and power control. *Ad Hoc Networks*, *186*, 104180. https://doi.org/10.1016/j.adhoc.2026.104180

[18] Jiao, L., Shao, Y., Sun, L., Liu, F., Yang, S., Ma, W., Li, L., Liu, X., Hou, B., Zhang, X., Shang, R., Li, Y., Wang, S., Tang, X., & Guo, Y. (2024). Advanced Deep Learning Models for 6G: Overview, Opportunities, and Challenges. *IEEE Access*, *12*, 133245-133314. https://doi.org/10.1109/ACCESS.2024.3418900

[19] Kha, H. H., & Quyet, P. V. (2025). Multi-objective optimization for active IRS-aided multi-group multicast systems with energy harvesting, integrated sensing and communication. *Physical*

Communication, *69*, 102549. https://doi.org/10.1016/j.phycom.2024.102549

[20] Khan, A., Mohsan, S. A. H., Elfikky, A., Boghdady, A. I., Ahmad, S., & Innab, N. (2025). A survey of intelligent reflecting surfaces: Performance analysis, extensions, potential challenges, and open research issues. *Vehicular Communications*, *51*, 100859. https://doi.org/10.1016/j.vehcom.2024.100859

[21] Letaief, K. B., Shi, Y., Lu, J., & Lu, J. (2022). Edge Artificial Intelligence for 6G: Vision, Enabling Technologies, and Applications. *IEEE Journal on Selected Areas in Communications*, *40*(1), 5-36. https://doi.org/10.1109/JSAC.2021.3126076

[22] Lu, J., Shi, Z., Wang, Y., Pan, C., & Zhang, S. (2023). Multi-index evaluation learning-based computation offloading optimization for power internet of things. *Physical Communication*, *56*, 101949. https://doi.org/10.1016/j.phycom.2022.101949

[23] Naguib, K. M., Ibrahim, I. I., Elmessalawy, M. M., & Abdelhaleem, A. M. (2024). Optimizing data transmission in 6G software defined networks using deep reinforcement learning for next generation of virtual environments. *Scientific Reports*, *14*(1), 25695. https://doi.org/10.1038/s41598-024-75575-y

[24] Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., Niyato, D., Dobre, O., & Poor, H. V. (2022). 6G Internet of Things: A Comprehensive Survey. *IEEE Internet of Things Journal*, *9*(1), 359-383. https://doi.org/10.1109/JIOT.2021.3103320

[25] Noor, K., Imoize, A. L., Adedosu Adelabu, M., & Lee, C.-C. (2025). A Comprehensive Survey on AI-Assisted Multiple Access Enablers for 6G and beyond Wireless Networks. *CMES - Computer Modeling in Engineering and Sciences*, *145*(2), 1575-1664. https://doi.org/10.32604/cmes.2025.073200

[26] Qian, L. (2022). Joint Multi-domain Resource Allocation and Trajectory Optimization in UAV-assisted M-IoT Networks. *TechRxiv*(0105). https://doi.org/10.36227/techrxiv.17704463.v1

[27] Richa, Tyagi, A. K., & Nair, M. M. (2026). Chapter 5 - Artificial intelligence-enabled security systems for 6g wireless networks: Algorithms, strategies, and application. In J. B. Awotunde, A. K. Bhoi, P. Barsocchi, & V. H. C. de Albuquerque (Eds.), *Fundamentals of Fog Computing and the Internet of Things for Smart Healthcare* (pp. 87-100). Elsevier. https://doi.org/10.1016/B978-0-443-33439-9.00018-0

[28] Shafin, R., Liu, L., Chandrasekhar, V., Chen, H., Reed, J., & Zhang, J. C. (2020). Artificial Intelligence-Enabled Cellular Networks: A Critical Path to Beyond-5G and 6G. *IEEE Wireless Communications*, *27*(2), 212-217. https://doi.org/10.1109/MWC.001.1900323

[29] She, H., Yan, L., & Guo, Y. (2024). Efficient End–Edge–Cloud Task Offloading in 6G Networks Based on Multiagent Deep Reinforcement Learning. *IEEE Internet of Things Journal*, *11*(11), 20260-20270. https://doi.org/10.1109/JIOT.2024.3372614

[30] Shome, S., Das, S., Das, S., & Pal, D. (2025). An extensive review of THz communication in 6G: Facilitating technologies with edge computing and native AI. *Franklin Open*, *13*, 100434. https://doi.org/10.1016/j.fraope.2025.100434

[31] Singh, J., Bagga, S., & Kaur, R. (2026). Resource-efficient federated edge-AI over wireless edge networks for intelligent systems. *Physical Communication*, *76*, 103131. https://doi.org/10.1016/j.phycom.2026.103131

[32] Sittakul, V., Ioannou, I., Nagaradjane, P., & Vassiliou, V. (2025). Intelligent congestion control in 5G URLLC Software-Defined Networks using adaptive resource management via Reinforced Dueling Deep Q-Networks. *Journal of Network and Computer Applications*, *242*, 104276. https://doi.org/10.1016/j.jnca.2025.104276

[33] Sundaramoorthy, P., Parikh, M., Keshireddy, S. R., Penumatcha, A. K., & Jayabalan, K. (2025).

Adaptive Resource Management in Cloud-Native Databases: A Study of Autonomous Scaling and Query Optimization. 2025 10th International Conference on Communication and Electronics Systems (ICCES),

[34] Wang, X., Wu, P., Hu, Y., Cai, X., Song, Q., & Chen, H. (2023). Joint Trajectories and Resource Allocation Design for Multi-UAV-Assisted Wireless Power Transfer with Nonlinear Energy Harvesting. *Drones*, *7*(6), 354. https://doi.org/10.3390/drones7060354

[35] Xu, Y., Xu, H., Chen, X., Zhang, H., Chen, B., & Han, Z. (2025). Blockchain-Based AR Offloading in UAV-Enabled MEC Networks: A Trade-Off Between Energy Consumption and Rendering Latency. *IEEE Transactions on Vehicular Technology*, *74*(11), 18065-18080. https://doi.org/10.1109/TVT.2025.3581015

[36] Yamsani, N., & Chenna Reddy, P. (2026). EdgeSched-DQN: An intelligent deep reinforcement learning-based framework for optimized task scheduling in edge-cloud environments. *Array*, *29*, 100645. https://doi.org/10.1016/j.array.2025.100645

[37] Zhang, J., Yu, L., Liu, S., Cai, Y., Zhang, Y., Xing, H., & Jiang, T. (2025). Wireless Environmental Information Theory: A New Paradigm Toward 6G Online and Proactive Environment Intelligence Communication. *Engineering*. https://doi.org/10.1016/j.eng.2025.07.028

[38] Zhang, W., Zhan, D., Yu, H., Zhang, L., Zhao, B., Du, X., & Tian, Z. (2026). Distributed machine learning for next-generation communication networks: A survey on privacy, fairness, efficiency, and trade-offs. *Information Fusion*, *126*, 103657. https://doi.org/10.1016/j.inffus.2025.103657