



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 Saddam², Basim Abood³¹ Department of Communication Engineering, College of Engineering, University of Sumer, Dhiqar, Iraq² Communication Engineering, College of Engineering, University of Sumer, Dhiqar, Iraq³ Communication Engineering, College of Engineering, University of Sumer, Dhiqar, Iraq

ARTICLE INFO

Article history:

Received 12 December 2025

Received in revised form 15 May 2026

Accepted 9 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 network heterogeneity, high dynamism of traffic demands, channel conditions and diverse Quality-of-Service (QoS) requirements in Beyond Fifth-Generation (B5G) and Sixth-Generation (6G) heterogeneous wireless networks create new difficulties for autonomous resource allocation. Static and reactive methods are not adaptable enough, and current learning based methods typically do not have a built-in mechanism for predicting, multi-objective, and temporally stable resource decisions. To fill this, this paper introduces a hybrid Decision Support System (DSS) for autonomous multi-objective resource allocation with Multi-Objective Optimization (MOOP), Deep Reinforcement Learning (DRL), Predictive Environment Model (PEM) and Stability-Aware Optimization (SAO) under a single decision-making framework. The proposed DSS integrates available network observation data with prediction of environmental conditions and is able to automatically infer resource-allocation decisions based on multiple and possibly competing criteria. DRL enables adaptive policy learning, PEM enables proactively making decisions for the future network state, MOOP enables balancing throughput, latency, fairness and energy efficiency, and SAO enables temporal consistency between successive allocation decisions. The framework also considers Signal-to-Interference-plus-Noise Ratio (SINR) and data rate to determine the quality of communication-links. A large number of simulation experiments are performed to compare the proposed DSS with static, reactive and conventional DRL-based resource-allocation methods. It shows that the results are better in the following aspects: Communication efficiency, Resource utilization, Latency, Fairness, and Energy efficiency. Convergence and temporal analyses also show quicker and more regular decision behavior and the SINR and data-rate analyses confirm better communication-link quality under the assumed operational conditions. As a whole, the results indicated that adding AL, PEWA, multi-objective optimization, and stability-aware decision-making to a DSS is a good way to autonomously manage the use of resources in a dynamic B5G and 6G wireless heterogeneous network while maintaining a balance between the conflicting goals.

1. Introduction

Fifth generation (5G) and the upcoming beyond-fifth generation (B5G) and sixth generation (6G)

* Corresponding author.

E-mail address: jafaarfahad@gmail.com, j.fahad@uos.edu.iq<https://doi.org/10.5281/zenodo.23299292>

wireless networks have led to an ever-more heterogeneous communication environment with increasing dynamics. [1]. The combination of macro base stations, small cells, heterogeneous users, different services demands, and time-varying wireless environments brings significant challenges for autonomous management of the network resources. In these types of environments, the allocation of resources needs to account for the availability of the spectrum, the power with which it is transmitted, the association of users, the interference, the dynamics of traffic and the requirements for quality of service (QoS). Typical resource allocation methods are based on predetermined rules, static optimization or on-demand control techniques. [2]. While they may work well in relatively stable network environments, their performance may suffer in environments with dynamic traffic loads, channel states, interference, and mobility. The decision mechanisms needed to enable autonomous resource management must therefore be able to detect the current state of the network, predict future network conditions, assess competing decisions and choose the optimal resource allocation actions. [3]. These requirements are clearly used in heterogeneous wireless networks as a multi-criteria decision-making problem. Network controllers need to make decisions on how to allocate the resources to the competing users, considering throughput, latency, fairness, and energy efficiency. An improvement in one criterion can lead to a degradation in another criterion, e.g. increasing the transmission power may result in an increase in throughput but also in an increase in energy consumed and in interference. An effective autonomous resource allocation mechanism should not be optimized in isolation, rather it should enable coordinated decision making for multiple QoS and operational criteria. [4]. The main research problem in this work is the problem of the autonomous resource allocation decision making in the heterogeneous wireless networks with continuously changing network conditions and service requirements. The process of decision making includes the selection of resource allocation actions with power constraints, spectrum constraints, interference constraints and QoS constraints considering multiple and potentially conflicting performance objectives. Leverage the network controller to monitor the current heterogeneous wireless environment and to decide on an allocation action, including transmission power, bandwidth allocation, user association and scheduling. The selected action should contribute to the enhancement of the network utility, while preserving network QoS requirements and avoiding network resource allocation behavior that could be prone to instability. [5]. The challenge is not just to identify an optimal resource allocation solution at any given moment in time, but to make adaptive decisions at every moment in time in a manner that is uncertain. Thus, the proposed study is aimed at designing an autonomous wireless resource allocation problem as an intelligent multi-objective decision problem, where decision factors include the throughput, latency, fairness, and energy efficiency. Deep Reinforcement Learning (DRL), a Predictive Environment Model (PEM), Multi-Objective Optimization (MOOP) and Stability-Aware Optimization (SAO) are used in the decision process. The proposed system can be viewed from decision making point of view as Intelligent Decision Support System (IDSS) for the management of autonomous wireless networks. The network controller is an independent decision maker which monitors network conditions continuously, predicts future environmental conditions, evaluates various performance criteria, and decides on actions for resource allocation. Decision variables are transmission power, bandwidth allocation, user association and scheduling decisions. The decision criteria include throughput, latency, fairness, and energy efficiency, while the decision process is subject to resource, QoS, interference, and stability constraints. The proposed framework therefore provides decision support by combining current network observations with predicted future conditions and multi-objective optimization before selecting the final resource allocation action. In contrast to conventional optimization approaches that determine resource allocation based

primarily on the current network state, the proposed decision-support perspective integrates prediction and adaptive learning into the decision process. The Predictive Environment Model provides information about potential future network conditions, the DRL agent learns an adaptive allocation policy, the MOOP formulation represents competing decision criteria, and the Stability-Aware Optimization mechanism regulates temporal variations in the selected decisions. This integrated process enables the framework to support autonomous and context-aware resource allocation decisions in dynamic heterogeneous wireless networks as shown in figure 1 [6].

In this figure1, there is a multi-tier heterogeneous network (HetNet) architecture comprising macro cells and small cells that are able to provide wireless coverage to a dense user environment. Users of the network vary from heterogeneous Quality of Service (QoS) requirements, such as 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 and small cells provide increased capacity and spectral efficiency in high-density areas. The system runs in a dynamic and non-stationary system, where user distribution, traffic demand, and channel state are constantly changing and require intelligent and adaptive resource allocation strategies. While significant progress has been made in the field of AI-assisted resource allocation for heterogeneous wireless networks, a crucial challenge in this research area is to create and deploy a fully autonomous and intelligent resource management scheme that can adapt to the rapidly evolving network scenarios in real time, thereby maintaining stability and scalability of the system [11]. The existing techniques, particularly those based on deep reinforcement learning and optimization techniques, are inadequate to satisfy multi-dimensional quality of service (QoS) requirements in highly non-stationary and interference-rich environments. The existing models are trained in simplified and/or partially-stationary environments, and are not robust enough in the real world in which the services are required, with bursty traffic and mobile users.

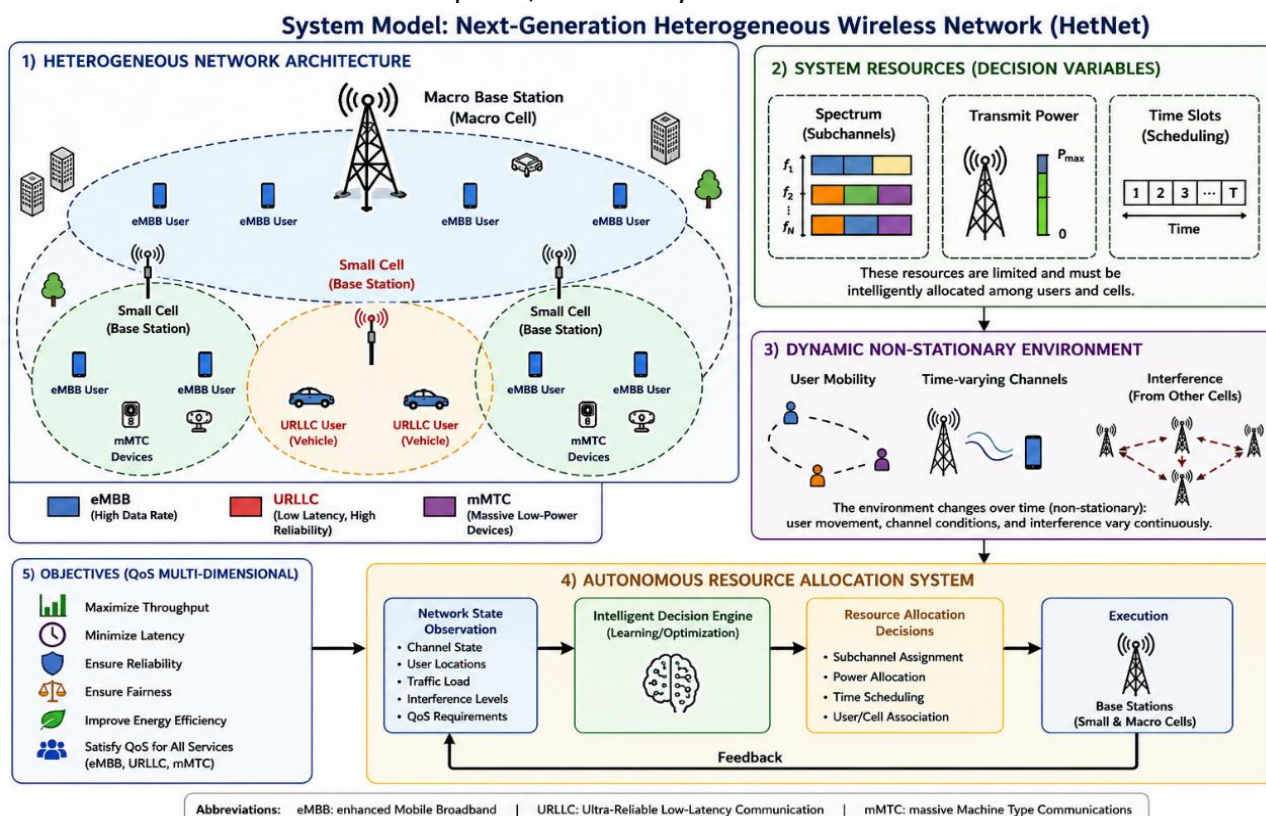


Fig.1: Architecture of a Multi-Layer Heterogeneous Wireless Network with Heterogeneous QoS Requirements

Moreover, there is still a lack of integration between the fast adaptive decision making mechanisms and guarantees of long-term stability, which means a compromise between responsiveness and reliability. It shows the necessity of a common framework which allows to easily install real-time adaptability, scalable architectures and QoS assurance in a coherent way regardless the network conditions, an open and challenging issue in the next generation wireless systems [12; 13].

Existing research has demonstrated the effectiveness of DRL, multi-objective optimization, predictive models, and QoS-aware resource allocation techniques in addressing individual aspects of wireless network management. However, several limitations remain when these approaches are considered from an integrated decision-making perspective. First, many existing resource allocation approaches focus primarily on optimizing one or a limited number of network performance indicators, which may result in an undesirable trade-off among throughput, latency, fairness, and energy efficiency. Second, conventional DRL-based allocation methods generally make decisions based on the observed network state without explicitly incorporating information about future environmental conditions [7]. Third, predictive resource management approaches can improve proactive decision-making but may not sufficiently address the stability of consecutive allocation decisions. Fourth, multi-objective approaches provide a mechanism for balancing competing criteria, but their integration with adaptive learning and predictive environmental information remains insufficiently explored as given in figure1. These limitations indicate the need for an integrated decision-support mechanism capable of combining adaptive learning, environmental prediction, multi-objective decision criteria, and temporal stability within a unified resource allocation process. Therefore, this study proposes a hybrid framework that integrates DRL, PEM, MOOP, and SAO to support autonomous resource allocation decisions in next-generation heterogeneous wireless systems. The main objective of this study is to develop an autonomous intelligent decision-support framework for multi-objective resource allocation in next-generation heterogeneous wireless networks [8]. To achieve this objective, the study pursues the following specific objectives: To formulate wireless resource allocation as a multi-criteria decision-making problem involving throughput, latency, fairness, and energy efficiency. To develop a DRL-based autonomous decision mechanism capable of adapting resource allocation decisions to dynamic network conditions. To add a Predictive Environment Model, generate information on future state of the network and support proactive resource allocation decisions. For incorporating Multi-Objective Optimization to satisfy conflicting decision parameters of QoS and energy. Add Stability-Aware Optimization to minimize unwanted variation between subsequent decisions on resource allocation. Compare the proposed framework with the traditional resource allocation scheme and DRL-based scheme under the various wireless network environments. To explore the implications of autonomous resource allocation for network operators and wireless network managers in a practical and managerial sense. Decision-oriented formulation: The study proposes an autonomous, wireless resource allocation problem as an intelligent, multi-criteria decision-making problem where the resources to be allocated include throughput, latency, fairness and energy efficiency, and the constraints comprise the network and the QoS constraints. Autonomous Resource Allocation Decision Support Framework: The proposed hybrid framework combines DRL, Predictive Environment Model (PEM), Multi-Objective Optimization (MOOP) and Stability-Aware Optimization (SAO) to provide support to autonomous resource allocation decisions in heterogeneous wireless networks. Predictive decision capability: The PEM is integrated into the decision process, and can provide the information about the future network condition, which allows the controller based on DRL to make more proactive than merely reactive resource allocation decisions. Multi-criteria

resource allocation: The proposed framework takes into account throughput, latency, fairness and energy efficiency, where conflicting objectives occur during the resource management of autonomous networks. STABILITY-AWARE DECISION MECHANISM (SAO): SAO is added to control temporal dynamic changes in the resource allocation decisions, while reducing the instability in the autonomous decision process. Comprehensive evaluation: The proposed framework is compared with the traditional allocation methods and DRL-based baselines using various network performance metrics such as prediction accuracy, convergence, SINR, data rate, latency, fairness and energy efficiency. Engineering management perspective: The study interprets autonomous resource allocation as a decision-support mechanism that can assist telecom operators and network managers in making adaptive resource management decisions under dynamic network conditions [7; 11].

A Multi-Objective Optimization Problem (MOOP) is formulated for the resource allocation and system optimization problem in the considered heterogeneous network (HetNet). This formulation is driven by the competing performance metrics that are necessary to be optimized in next generation wireless networks [16]. In particular, the optimization framework is designed to enhance multiple performance measures. First, the system throughput can be maximised for maximum overall spectral efficiency, and to enable high data-rate services such as enhanced Mobile Broadband (eMBB). Second, end-to-end latency is reduced to comply with the high requirement for delay, particularly for the applications in the Ultra-Reliable Low-Latency Communications (URLLC). Third, maximising energy efficiency, which is crucial for sustainability of network operation, by minimising the power consumption at base station and user equipment. Secondly, the formulation ensures that users are treated fairly, without the risk of resource starvation and a fair resource distribution amongst various user groups [17]. Finally, the model guarantees multi-level Quality of Service (QoS) requirements are met, and allocates network resources for each service type (eMBB, URLLC – ultra reliable low latency communication and mMTC – massive machine-type communication), depending on its specific performance requirements. The mathematics is treated as a single entity and not as separate entities, and this gives rise to a tradeoff analysis of goals that conflict, e.g. throughput vs energy, latency vs fairness. Several practical constraints that are based on the real world constraints imposed on wireless networks in the optimization problem. A few of these are spectrum scarcity, spectrum reuse in dense heterogeneous deployments (inter-channel interference); and heterogeneous QoS constraints, each service type with different QoS requirements such as minimum data rate, maximum latency/reliability etc. The formulated MOOP can be solved by advanced solution methods such as machine learning algorithms to optimize resource allocation and evolutionary optimization methods, as a whole, it is a compromise between the performance optimization and physical network constraints {Gülmez, 2026 #40}[13].

The proposed framework has an impact on multiple stakeholders of the operation and management of wireless networks. The decision-support mechanism can be used to support the dynamic utilization of resources and QoS management by a telecom operator. Network managers can employ the framework to balance competing operational objectives, including capacity, latency, fairness, and energy consumption. Infrastructure planners can benefit from predictive information when evaluating resource demand and network conditions. From an engineering management perspective, the framework provides an autonomous mechanism for supporting operational decisions in heterogeneous wireless networks rather than relying exclusively on manually configured or static allocation policies. 9 concludes the study [14; 15].

Deep Reinforcement Learning (DRL) has become an important approach for autonomous resource management in heterogeneous and next-generation wireless networks because it can learn

resource allocation policies through continuous interaction with dynamic environments. Alwarafy et al. [1] provided a comprehensive survey of DRL-based radio resource allocation and management in next-generation heterogeneous wireless networks, highlighting the ability of DRL to address complex resource management problems without requiring an explicit analytical solution for every network condition. However, as a survey study, the work does not provide an integrated resource allocation mechanism that jointly combines prediction, multi-objective optimization, and stability-aware control. Diamanti et al. [6] investigated an energy-efficient rate-splitting multiple access framework based on DRL. Their study demonstrates the potential of learning-based resource management for improving energy efficiency while considering communication performance. However, the focus is on a specific multiple-access scenario rather than a general heterogeneous network with simultaneously conflicting QoS criteria. Naguib et al. [23] investigated deep reinforcement learning for data transmission optimization in 6G software-defined networks. Their work demonstrates the applicability of DRL to adaptive network resource management in emerging 6G environments. Nevertheless, the study primarily emphasizes transmission optimization and does not explicitly integrate a predictive environmental model with a stability-aware multi-objective decision mechanism. She et al. [30] investigated intelligent congestion control in 5G URLLC software-defined networks using adaptive resource management based on a reinforced dueling deep Q-network. Their work is particularly relevant to QoS-aware resource management because it addresses congestion and resource adaptation for latency-sensitive services. Nevertheless, the focus is primarily on congestion control and URLLC-oriented resource management rather than a unified multi-objective framework that combines prediction, adaptive learning, and temporal stability. Yamsani and Chenna Reddy[37] proposed EdgeSched-DQN for intelligent task scheduling in edge-cloud environments. Their study further demonstrates the applicability of DQN-based learning to dynamic scheduling problems. However, the problem domain is edge-cloud task scheduling rather than heterogeneous wireless resource allocation, and the framework does not explicitly address the simultaneous optimization of wireless throughput, latency, fairness, and energy efficiency. Overall, these studies demonstrate that DRL can provide adaptive decisions in dynamic wireless and edge environments. However, the existing approaches represented in the present reference set generally address specific resource management problems or emphasize selected performance objectives. This motivates the development of a more integrated decision framework capable of jointly considering multiple QoS criteria, future network conditions, and temporal stability.

Wireless resource allocation is inherently a multi-objective problem because network operators must balance several competing performance criteria. Kha and Quyet [20] investigated multi-objective optimization for active IRS-aided multi-group multicast systems with energy harvesting and integrated sensing and communication, demonstrating the importance of simultaneously considering multiple system objectives in emerging wireless architectures. However, their study focuses on an active IRS-assisted multicast scenario rather than autonomous resource management in heterogeneous cellular networks. Benatia et al.[4] . investigated energy savings and coverage optimization using a multi-objective evolutionary approach for edge WiFi sensor deployment. The study illustrates how multiple conflicting objectives can be addressed through evolutionary optimization. Nevertheless, the considered problem is deployment-oriented and does not incorporate adaptive DRL-based decision-making or predictive network-state information. Etheredge et al[9] . proposed a multi-objective scheduling approach for service-based applications in multi-cluster cloud-edge-IoT infrastructures. Their work demonstrates the importance of balancing multiple objectives in complex resource scheduling environments. However, the problem is primarily concerned with cloud-edge-IoT service scheduling rather than wireless resource

allocation under channel, interference, and QoS constraints. Lu et al[23] . investigated multi-index evaluation learning-based computation offloading optimization for power Internet of Things. Their work highlights the importance of considering multiple performance indicators in resource-related decisions. Nevertheless, the emphasis is on computation offloading rather than joint wireless resource allocation involving power, bandwidth, user association, throughput, latency, fairness, and energy efficiency. These studies confirm the suitability of multi-objective optimization for complex engineering resource-management problems. However, multi-objective optimization alone does not provide an adaptive mechanism for continuously learning resource allocation policies from changing wireless environments. Therefore, MOOP is incorporated in the proposed framework as the decision criterion layer, while DRL provides adaptive decision-making and the predictive and stability mechanisms address future environmental variation and temporal consistency. Reactive resource allocation makes decisions primarily from the current network state and may therefore respond only after congestion, channel degradation, or QoS deterioration has occurred. Predictive resource management provides an alternative by incorporating information about potential future network conditions into the resource allocation process. Dutta et al. [7] investigated predictive optimal channel allocation in integrated sensing, communication, and computation systems. Their work demonstrates the value of prediction in improving resource allocation decisions before future network conditions are realized. However, the considered framework is specific to integrated sensing, communication, and computation environments and does not combine predictive allocation with DRL-based learning and stability-aware control in a heterogeneous cellular network. {Noor, 2025 #26@@author-year} examined edge-AI resource recommendation techniques for 5G/6G channels, highlighting the increasing role of AI-enabled resource recommendation in future wireless environments. The study provides a broader perspective on intelligent resource management but does not formulate the complete decision process as an integrated DRL–prediction–stability framework. Zhang[38] discussed wireless environmental information theory as a paradigm toward online and proactive environmental intelligence communication for 6G. This work supports the broader concept that future wireless systems should move from reactive communication mechanisms toward proactive environmental awareness. However, the study does not specifically provide the resource allocation architecture proposed in the present work. Ejaz et al[8] . proposed FAIR-DQL for fairness-aware deep Q-learning in resource allocation and RIS optimization for high-altitude platform networks. Their work is particularly relevant to the fairness dimension of autonomous resource management. Nevertheless, the considered HAP/RIS environment differs from the heterogeneous cellular network considered here, and the framework does not explicitly integrate predictive environmental modeling and stability-aware optimization. Consequently, the existing studies support the importance of proactive and predictive intelligence but leave an opportunity for integrating prediction directly into a multi-objective DRL-based resource allocation decision loop. The increasing heterogeneity of 5G and 6G services has made QoS-aware resource management an essential requirement. Hamdi et al. [14] investigated QoS-aware network slicing and resource management for Internet of Vehicles in 5G networks. Their study emphasizes the importance of incorporating service requirements into resource management decisions. However, the application is mainly oriented toward network slicing and vehicular communication rather than a general heterogeneous wireless resource allocation framework. Sittakul et al[33] . Dealt with intelligent congestion control for 5G URLLC software-defined network with adaptive resource management. The study is applicable to the current research due to the need for making resource allocation decisions in response to network congestion and QoS conditions. The proposed approach, however, is more oriented towards congestion control and not in balancing

throughput, latency, fairness and energy efficiency at the same time. Islam et al. [17] studied resource sharing and power control in underlay many-to-many vehicular communication that takes into account QoS. They have illustrated the need for conjunctive resource and power control for QoS improvement in their study work. However, the multi-tier, heterogeneous network, which was considered in this study, is different from the vehicular communication scenario considered here. The authors of [18] analyzed the state of the art in deep learning for 6G and elaborated on opportunities and challenges that arise with intelligent wireless systems. Likewise, Islam and Noh [18] summarized the performance evaluation and optimization of 6G from the point of view of KPIs, tools and AI models. These studies focus on the shift towards AI-driven and performance-conscious 6G networks, excluding the integrated autonomous resource allocation mechanism introduced in this study. All these studies confirm that it is necessary to include the requirements for QoS in intelligent network management. Still, existing work lacks a single mechanism that can learn allocation policies, predict future conditions, balance multiple QoS criteria, and control the temporal instability. Two important criteria that can conflict with the goal of throughput maximization when allocating wireless resources are energy efficiency and fairness. Diamanti et al. [6] illustrated the application of DRL in the field of energy-efficient wireless communication and have shown that learning can be used to optimize communication performance while taking into account energy consumption. In addition to coverage optimization, energy savings were also taken into account by Benatia et al. [4] in a multi-objective evolutionary approach. Lu et al. [23] studied multi-index evaluation learning for computation offloading, showing that multiple performance indicators are significant for the resource-management decisions. Ejaz et al. [8] particularly designed a deep Q learning method for resource allocation and optimizing RISs to incorporate fairness. The results of these studies show the value of energy efficiency and energy equity as resource-management goals. There exists, however, a need for a treatment of these criteria individually or in specific contexts that attempts to address the trade-offs among throughput, latency, fairness, and energy efficiency in dynamic heterogeneous wireless networks. The proposed framework therefore incorporates these criteria into a unified multi-objective decision formulation. Adaptive learning introduces an important challenge in dynamic wireless environments: resource allocation decisions may change substantially between consecutive decision intervals because of changing channel conditions, traffic demand, user mobility, and learning behavior. Therefore, achieving high instantaneous performance is not sufficient; the decision process should also maintain temporal consistency and stable network operation. Chen et al [5] . discussed resource allocation in software-defined networks and identified important research challenges associated with intelligent resource management. Gadouh and Sellami [11] investigated machine-learning-based resource allocation and dynamic placement for virtualized network functions, illustrating the importance of adaptive resource decisions in dynamic network environments. Heydari et al. [16] considered a hierarchical deep-learning and distributionally robust framework, providing further evidence of the importance of robustness when intelligent optimization is applied to dynamic engineering systems such as shown in table 1. The remainder of this paper is organized as follows. Section 2 presents a critical review of existing studies related to DRL-based resource allocation, multi-objective optimization, predictive resource management, stability-aware optimization, and intelligent decision-support approaches. Section 3 describes the heterogeneous wireless system model and formulates the resource allocation problem from a multi-criteria decision-making perspective. Section 4 presents the proposed autonomous decision-support framework, including the DRL decision engine, Predictive Environment Model, Multi-Objective Optimization, and Stability-Aware Optimization mechanisms.

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. [6]	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. [7]	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. [30]	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. [24]	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.	Hamdi et al. [14]	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, energy efficiency, and throughput	Supports QoS-aware resource allocation
7.	Sittakul et al. [33]	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

Table 1 (cont...)

Critical comparison 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
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 [20]	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	Etheredge et al. [9]	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.	Ejaz 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. [26]	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.	Yamsani and Chenna Reddy [37]	6G	Proactive environmental intelligence	Environmental awareness	Promotes proactive environmental intelligence for future 6G communication	Does not provide a complete resource allocation mechanism or integrate environmental prediction with DRL and stability control	Motivates the 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

Section 5 describes the simulation environment, experimental configuration, baseline methods, and evaluation methodology. Section 6 presents and discusses the simulation results. Section 7 discusses the managerial and practical implications of the proposed framework. Section 8 presents the limitations and future research directions. Finally, Section.

Although these studies support the need for adaptive and robust intelligent decision mechanisms, the specific stability-aware optimization mechanism developed in the present study is designed as an additional layer for regulating consecutive wireless resource allocation decisions. Thus, the role of stability in the proposed framework is not treated as a replacement for DRL but as a complementary mechanism that constrains abrupt allocation changes while maintaining the multi-objective performance requirements [16]. Based on the critical comparison, four significant literatures gaps are identified. Second, although strong adaptive decision-making ability are shown by existing DRL-based approaches, many of them are merely focused on application-specific problems like task offloading, congestion control, vehicular communication, or energy-efficient wireless transmission. Second, multi-objective approaches are able to effectively express trade-offs between different objectives, but they effectively do not incorporate continuous DRL-based policy learning, along with future-state prediction. Third, predictive resource management can offer proactive decision-making capabilities, however, the proactive prediction isn't always utilized in a closed-loop wireless resource allocation mechanism that also takes decision stability into account. Fourth, the presented methods consider some individual measures such as QoS-aware and fairness-aware methods, but, with respect to the dynamic network environment, it is still not enough to consider all the measures together including throughput, latency, fairness, and energy efficiency. [14]. These observations motivate the development of an integrated autonomous decision-support framework in which DRL provides adaptive policy learning, PEM provides future-state awareness, MOOP represents conflicting decision criteria, and SAO regulates the temporal stability of resource allocation decisions. The proposed framework therefore addresses the identified gaps by treating wireless resource allocation as a continuous multi-criteria decision-making problem rather than as an isolated optimization task.

The proposed study differs from the reviewed approaches by integrating four complementary capabilities within a single decision loop. DRL provides adaptive policy learning from network interactions; PEM introduces proactive information about future network conditions; MOOP establishes the multi-criteria decision structure; and SAO regulates the temporal evolution of resource allocation decisions [18]. The resulting framework is therefore positioned as an intelligent decision-support mechanism for autonomous resource management in heterogeneous 5G/6G networks. The comparative analysis in Table 1 indicates that existing studies have addressed important individual dimensions of intelligent wireless resource management. DRL-based studies have demonstrated the capability of adaptive learning for resource allocation, transmission optimization, and fairness-aware decisions [1; 6; 9; 29; 32; 35]. Other studies have focused on multi-objective optimization to address competing performance requirements such as energy consumption, coverage, sensing, and communication performance [4; 8; 19]. Predictive and proactive approaches have further demonstrated the importance of incorporating information about future network conditions into resource management decisions [7; 27; 37]. However, the reviewed studies generally emphasize one or two of these capabilities rather than integrating them into a unified autonomous decision-making process. In particular, the combination of adaptive DRL-based decision-making, predictive environmental information, multi-objective QoS optimization, and temporal stability remains insufficiently addressed. In addition, there are a number of techniques published that are tailored for a certain application scenario such as vehicular networks,

RIS-supported systems, HAP networks, edge-cloud systems, or task offloading, restricting their direct use in general heterogeneous cellular networks. Thus, the proposed study aims at filling this gap by combining DRL, PEM, MOOP and SAO in a single intelligent decision support approach for B5G/6G heterogeneous wireless networks. The framework sees resource allocation as a continuous multi-criteria decision making problem that considers throughput, latency, and fairness and energy efficiency with network and QoS constraints.

2. DESCRIPTION SYSTEM

The system model envisioned in the proposed framework is meant to be a realistic one for a next generation of a highly dynamic and uncertain heterogeneous wireless network (HetNet) [18]. The autonomous resource allocation problem is formulated and solved with the aid of the model. The network architecture is composed of a series of hierarchical layers, often with macro cells for wide area coverage and small cells to increase capacity in dense urban areas. Such a multi-layered architecture will play a fundamental role in 5G and beyond networks, both for the goal of coverage extension and of capacity densification. In such an environment, a lot of mobile users share the same physical resource and make different traffic demands for their applications due to the diversity of their demands. The users are classified into three categories: enhanced Mobile Broadband (eMBB), having high data rate requirements; Ultra-Reliable Low-Latency Communication (URLLC) with very low latency and high reliability requirements; and massive Machine Type Communications (mMTC) with extremely large number of low-power devices sending small amount of data. The network is modeled as a non-stationary dynamic system, to accurately reflect the real world behavior, where conditions change over time. These include the shifts in user mobility, channel fading and shadowing effects of wireless channels, and the interference resulting from simultaneous transmissions in overlapping cells. Thus, the environment is not predictable and static, thereby making resource allocation decisions more complex. In this system, critical communication resources are clearly identified as variables that can be controlled. These are: radio spectrum (frequency bands or sub-channels), transmission power levels of each user or base station and the assignment of time slots for data transmission. These resources are limited, and should be allocated wisely between users and between cells to optimize the use of the network [19]. The main issue of the system model is therefore how to formulate resource allocation as a real-time decision-making problem, that is, the system needs to continuously monitor the state of the network and adjust the distribution of resources continuously. The goal is to achieve the best possible overall network performance, meet varying QoS demands, ensure fairness of the network to all users and maximize efficient use of limited wireless resources in a highly dynamic and interference rich environment.

2.1 System Overview and Modeling Perspective

The considered heterogeneous network (HetNet) is modeled as a multi-tier wireless system, which is composed of a dense set of users with heterogeneous service requirements, and a set of macro base stations and small cells that collectively serve them. The system is under dynamic channel conditions and varying user distribution, traffic demand and interference levels over time. In order to mathematically describe this system, we introduce:

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

The system is inherently cross-layer coupled, meaning that physical layer variables (e.g., SINR and power) directly affect network layer metrics (throughput, latency, fairness, and energy

efficiency).

User Association Model (Network Layer Coupling)

Each user must be connected to exactly one base station. This is modeled using a binary association variable.

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

- $x_{u,b} = 1$, if user u is associated with base station b
- $x_{u,b} = 0$, otherwise

The uniqueness constraint ensures stable network connectivity:

$$\sum_{b \in B} x_{u,b} = 1 \quad u \in U \quad (1)$$

This constraint gives rise to a combinatorial structure and the problem becomes NP-hard since the user association needs to be optimized along with power and resource allocation [20]. Physical Layer Model Signal-to-Interference-plus-Noise Ratio (SINR). Signal-to-Interference-plus-Noise Ratio (SINR): It is an important metric in wireless communication systems that quantifies the quality of the received signal by dividing the power of the useful signal by the sum of the interference and noise power. It is mathematically expressed as:

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

- 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 additive white Gaussian noise (AWGN) power

A Multi-Objective Optimization Problem (MOOP) is a mathematical problem that deals with optimization of multiple performance objectives, which are often conflicting and usually have to be optimized simultaneously [21; 22]. Classical optimization typically has one objective function, e.g. maximizing throughput or minimizing cost. MOOP is commonly employed in wireless communication systems due to the fact that there are several interrelated factors that affect the performance of a communication system. A system can be designed for, say, maximizing throughput as well as minimizing latency, having enhanced energy efficiency and equitable access for users. Of course, these goals cannot co-exist without conflict because an improvement in one goal actually harms another goal. For example, higher transmission power results in higher throughput but lower energy efficiency and higher interference, whereas a strict fairness setting can cause the overall system capacity of the system to decrease. Due to these conflicts, it is difficult to find a single best solution – one that meets all objectives as well. However, MOOP results in a set of optimal trade-off solutions called Pareto optimal set. A solution is Pareto optimal if there are no objectives for which it can be improved without harming at least one other objective. Several solution methodologies have been used for solving MOOP such as weighted sum methods, ϵ -constrained methods, evolutionary algorithms (e.g., NSGA-II) and advanced machine learning algorithms (e.g., deep reinforcement learning) [23-25]. The methods attempt to approximate the "best possible" set of trade-offs, the Pareto front, between the conflicting goals. To conclude, MOOP is a rich methodology for the modeling and solution of complex engineering problems particularly when dealing with multiple performance measures that need to be optimized in modern wireless networks with practical system constraints. In MOOP, however, there are several objective functions, and they all

must be taken into account. In the mathematical terms it is written as:

$$\max F(x) = (F_1(x) + F_2(x) \dots F_K(x)) \quad (3)$$

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

$$\max J(x) = \sum_{k=1}^K w_K F_K(x) \quad (4)$$

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

$$P_i \leq P_i^{max} \quad A_i, \quad \sum_{u \in U} B_u \leq B_{total}, \quad QoS_u \geq QoS_u^{min} \quad A_u$$

These constraints guarantee power budgets are met, bandwidth is available and Quality of Service (QoS) requirements are met. However, due to the conflicting objectives, there is no solution which optimizes all the metrics. The system, however, tries to find Pareto optimal solutions – solutions which can be improved in one objective without worsening another {Qian, 2022 #25}{Noor, 2025 #26}.

2.2 Deep Reinforcement Learning (DRL) Layer

Deep Reinforcement Learning (DRL) layer is the heart of the proposed system, each base station is independently self-made decision making unit, which adapts to continuously changing network conditions. The DRL agent is not hard-coded with any rules or fixed optimization models, but instead, learns an optimal behavior with interactions with the wireless infrastructure over time [29]. In each time step, the agent observes the current network state, including the overall network conditions including traffic load, user distribution, interference and channel quality. The agent chooses an action that is related to a resource allocation decision, which can be user scheduling, bandwidth allocation, or transmission power adjustment. The environment will then elicit a response that will take the environment to a new state and provide a reward signal according to the quality of action. Learning is oriented to maximizing the total reward maximally over time rather than just in the present. This allows the agent to plan long-term strategies which maximize multiple performance measures such as throughput, latency, energy efficiency and fairness. The reward function is therefore carefully designed to capture these goals in a unified form, so that the agent will be rewarded with an overall optimal trade-off, instead of a single metric. Given the relation between the network state and the optimal action to take is very complex and nonlinear, the underlying policy or value function is approximated by deep neural networks. These networks allow the agent to process all inputs with high dimensional and provide generalization for various scenarios of the network. One of the important features of the DRL layer is that it is able to work in dynamic and non-stationary environments. In wireless networks, user mobility, fading of channels and interference patterns are constantly changing which requires the agent to continually adapt its policy. This is done through learning loops: experience from the past is used to improve decisions for the future. This process involves the agent taking a mixture of exploration and exploitation strategies – in other words, sometimes exploring to find better strategies and sometimes exploiting the knowledge it has gained. In a multi-cell heterogeneous network, the DRL agents are usually each contained in a base station, and all the agents run concurrently. While they do not normally interact directly, their decision making is connected implicitly because the action that is taken by one agent

will impact the environment that others will observe. This leads to a complex learning scenario as the environment seems non-stationary to each separate agent. Even with this difficulty, the DRL framework is effective because it continually revises the policy based on the changes witnessed. In summary, the DRL layer deals with a sequence of decisions, where the network's performance can be optimized based on past experiences. It offers flexibility and adaptability needed for modern wireless systems, for which all the dynamics do not need to be explicitly modelled with mathematical models that can be optimized on the fly [30; 31]. The intelligent decision making engine of the framework is the Deep Reinforcement Learning (DRL) Layer. Each base station works as an independent entity that learns the best way to allocate resources through continual exchange with the wireless environment. The agent gets the current state of the network as its observation at time step t :

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

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:

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

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

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

Where Tt is throughput, Lt is latency, Et is energy efficiency, and Ft is fairness, α, B, δ , and η balancing coefficients. The policy is approximated using deep neural networks:

$$\pi_{\theta}(\alpha t, St)$$

Where θ are the trainable parameters in the network. This learning process allows the agent to develop the ability to adapt to a dynamic and nonlinear wireless environment without having any explicit analytical models.

2.3 Predictive Environment Model

One important component of the autonomous resource allocation approach is the Predictive Environment Model (PEM), which predicts the network state, and not just the current state of the network [32]. The environment of a heterogeneous wireless system is naturally non-stationary in many aspects (e.g., user positions, channel quality, traffic demand, interference patterns, etc.) and the system is usually very dynamic as well. Based on local state only, may lead to a short sighted decision making, this is especially true in case of high mobility and bursty traffic. Hence, by introducing a predictive mechanism, the system can adapt its resource allocation approach in advance so that any performance degradation can be avoided. It is based on learning temporal patterns from a set of network historical and real-time data. These data generally consist of user mobility paths, SINR changes, traffic arrival rates and handover events. The predictive module can use time-series forecasting methods or lightweight supervised learning methods to predict network states for a short prediction horizon in the near future. These predictions can be for expected

channel conditions for each user, the expected congestion levels in specific cells, and the expected interference fluctuations caused by adjacent transmissions. The addition of predicted information to the decision-making process greatly improves the stability of the entire system. The resource allocation agent can not only consider the present situation, but also the future situation, allowing for more stable and forward-looking actions. This is especially relevant for applications that have strict latency or reliability requirements, and are subject to reaction time delays that may result in the violation of these requirements. The other significant advantage of a predictive model is that it can be used to increase the stability of algorithms that rely on learning, in particular reinforcement learning. In non-stationary environments, there can be sudden changes in the distributions of states that can result in unstable learning, slower convergence rates. These variations are smoothed with short term prediction, thus reducing uncertainties and providing a more consistent input for the learning agent. This results in more accurate policy updates and improved transfer of policies from one network situation to another. Also, it is to be efficient in computation which is to be applicable in real time. The model usually uses simplified or lightweight predictors that are not dependent on complex or heavy deep learning architectures, but rather achieve a balance between accuracy and latency. This design is crucial for large scale deployments where multiple base stations are required to function at the same time with limited computation resources [33; 34].

Predictive Environment Model (PEM): A computational system, which learns, estimates or simulates the evolution of an environment over time, based on action, external conditions or internal dynamics. The primary goal is to be able to predict future states of a system, and making decisions based on those predictions. A predictive environment model is a virtual model of the environment in engineering and intelligent systems that represents the real world. The system does not simply respond to what is observed at the moment but instead is used to anticipate what is likely to occur next in the observed process. This is particularly useful in autonomous systems, wireless networks, robotics and resource management. From a conceptual point of view, the model can be written as:

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

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 $f(\dots)$ is learned or designed predictive.

2.4 Stability-Aware Optimization

The primary goal of the Stability-Aware Optimization Module is to adapt the decisions made by the learning system in a way that is stable, smooth and reliable in a time-varying wireless network, particularly in a very dynamic wireless setting where conditions change quickly [10]. When operating in these conditions, the performance of a learning-based controller, like a deep reinforcement learning agent, can vary a lot from one time step to the next as a result of exploration, random changes in the channel, or abrupt shifts in user requests. The abrupt changes can cause negative effects on the system's performance such as the rise of signaling overhead, instability in user experience and inability to adhere to the QoS. The task of this module is thus to enable a controlled mechanism which controls the evolution of decisions over time and yet maintains good performance. The fundamental idea of this module is to think about the resource allocation as a problem that is coupled in time – decisions made today are to be similar to those made yesterday, not just some optimization problem solved at one point in time. From a mathematical point of view, it would be done by imposing restrictions on the evolution of the decision vector over time. Assume at time t the decision vector is $\{x_t\} = (x_{t1}, x_{t2}, \dots, x_{tK})$ of power allocation, bandwidth assignment and user association, respectively, then the stability would be enforced by limiting the difference

between two consecutive decisions. This is a restriction that ensures that the system will change smoothly and without any sudden changes or discontinuities. We do not want to jump from a large number of changes in the configuration, but we do want it to change slowly in a reasonable range that is dependent on the previous configuration. Stability is also directly incorporated in the objective function, through a regularization mechanism and also hard constraints. The original optimization goal, which usually has a goal of maximizing the network utility (for example, throughput, energy efficiency, fairness), is combined with a penalty term that dis-incentivizes making large deviations from previous decisions. Thus, the optimization problem is transformed to a balance between performance and stability, searching for the amount of influence the system has to place on the smooth behaviour in order to achieve faster performance gains. Obviously, therefore, it would be advantageous to the optimizer if he could keep to incremental changes only if they result in a significant improvement in the objective. A more stringent definition of stability is provided by the use of the Lyapunov theory, a formal method of investigating the behaviour over time of dynamic systems. A Lyapunov function is used for this purpose and it measures the deviation from an operating point of stable operation. This change in this function is called the Lyapunov drift, and the system is driven to a stable equilibrium with minimal change in this function while achieving the optimum performance objectives. This results in a drift-plus-penalty formulation that will attempt to stabilize the system and optimize the network efficiency. This formulation will prevent the system from oscillating too much, or diverging, even when subject to time varying conditions. Another important factor of stability in wireless networks is related to queue dynamics, especially for delay sensitive applications. The packets of data received to the system should be serviced at a rate that does not allow queues to become infinitely large. The requirement is modeled by a set of equations called queue evolution equations, which are used to determine the length of a queue at a given point in time, based on the number of customers that arrive and the length of time it takes to serve each of them. In this sense stability implies that the expected length of the queue is finite for all time, directly implying bounded delay. The system enables the system to enforce queue stability constraints, so that the Quality of Service requirements with particular emphasis on low-latency communications are always met. The module also addresses uncertainty in the system, including prediction errors for channel estimates, or variations in interference that are not expected. The formulation can be expanded to address worst case conditions in a set of uncertainty bounds other than optimizing for a single deterministic scenario. The Stability-Aware Optimization Module guarantees that decisions are smooth, reliable and temporally consistent over time. In learning-based systems, sudden jumps in control actions can cause higher signaling overhead and result in a poor user experience [28; 35]. To avoid this, a temporal regularization technique is used for optimization. The following decision smoothness constraint is imposed:

$$\|x_t - x_{t-1}\| \leq \epsilon$$

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

$$J_{stable}(x_t) = J(x_t) - \lambda \|x_t - x_{t-1}\|^2 \quad (10)$$

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

$$L(Q_t) = \frac{1}{2} Q^2 \quad (11)$$

The corresponding drift is:

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

The optimization minimizes the drift-plus-penalty term:

$$\min \Delta(Q_t) - V E[rt|Q_t] \quad (13)$$

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

$$Q_{t+1} = \max(Q_t - u_t, 0) + A_t \quad (14)$$

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

It is a very effective optimization method which ensures the performance of the system regardless of the environment in which the system operates compared to the one that is optimized. This results in the model having greater transferability and being less sensitive to inaccuracies in the model. The Stability-Aware Optimization Module's overall purpose is to "tune" the decisions made by the learning component. It renders the resource allocation a controlled dynamic system, with the addition of performance optimisation into temporal consistency and robustness. In order to be deployed practically, it is necessary to achieve the high efficiency of the system and reliable and stable operation in the long term [36; 37].

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 treating resource allocation as an isolated optimization task, the proposed framework establishes a decision process in which the network state is continuously observed, relevant network information is evaluated, and resource allocation actions are selected according to multiple and potentially conflicting performance requirements. The proposed decision process explicitly identifies the main elements of the resource allocation decision. The decision variables consist of transmission power, bandwidth allocation, user association, and scheduling decisions. The decision criteria include throughput, latency, fairness, energy efficiency, and Quality of Service (QoS), which are jointly considered because improvements in one criterion may adversely affect another. For instance, power up your wireless transmissions and boost received signal quality and throughput, but you will also boost received interference and energy inefficiency. On the other hand, if the fairness requirements are tightened, it can impact on the network throughput that can be achieved at the aggregate level. Therefore, the resource allocation issue must have explicit mechanisms that assess and cope with various objectives instead of optimizing a single performance measure. The base stations are linked to the Deep Reinforcement Learning (DRL) decision-making unit of the decision-maker. The agent monitors the network state, such as SINR, queue status, user distribution, interference state, and traffic demand at each decision interval and makes an optimal choice of resource allocation action. The Multi-Objective Optimization Problem (MOOP) is responsible for the multi-criteria formulation, and it encapsulates the trade-offs among the considered network objectives, while the DRL layer is responsible for the learning mechanism of the decision of the allocation of the resources based on the interaction with the wireless environment. The decision process is also assisted by the Predictive Environment Model (PEM), which is based on historical and current information from the network such as user mobility paths, SINR variations, traffic arrival rates and handover events and predicts the evolution of the network environment. The anticipated information is used in conjunction with the current network state to help the resource allocation mechanism plan for expected changes in the network state instead of only the instantaneous state. This predictive capability is especially important in the considered non-stationary environment, where user mobility, channel variations, traffic fluctuations and interference can significantly impact the success of a previously selected allocation decision.

The Temporal Consistency (TC) element is added to the decision process by the Stability-Aware Optimization (SAO) component. The chosen resource allocation decision is compared against the previous allocation state and too many changes are prevented by applying temporal regularization and stability constraints. The framework also takes into account the evolution of queues to enable stable operation and a QoS requirement. Consequently, the decision process does not seek only a high instantaneous network utility; it also considers the temporal behavior and stability of consecutive resource allocation decisions [38; 39].

3 SIMULATION ENVIRONMENT

The simulation parameters were carefully selected to provide a realistic, yet computationally efficient 5G heterogeneous network environment for the evaluation of a proposed Multi-Objective Optimization Problem (MOOP) framework. To demonstrate the typical urban deployment scenarios with coexistence of macro base stations, small cells, and mobile users under overlapping coverage scenarios, a 1000 x 1000 m² simulation area was used. Each base station is located at a larger scale, and the twelve small cells offer a reasonable compromise in terms of network coverage and capacity, along with meaningful load distribution and interference among cells. For this purpose, a total of 200 users were assumed to generate moderate to high network traffic, thus providing enough competition for radio resources and allowing user association, resource allocation and fairness to be evaluated in dynamic environments. The frequency of 3.5 GHz was chosen due to the widespread adoption of 5G New Radio (NR) configurations, and the system bandwidth of 100 MHz was chosen as it is considered a reasonable balance between coverage and spectral efficiency and the expected throughput. The simulation accounts for the 3GPP path loss model for Urban Macro (UMa), Rayleigh fading, and 8 dB log-normal shadowing, and thus can be made representative for the large-scale attenuation, multipath propagation, and obstructions due to the surrounding environment that are typical in urban environments. The user mobility is modeled by Random Waypoint model with speed varying from 3 to 30 km/h, which means the model is suitable for pedestrians and slow moving vehicles, and the channel quality, interference and handover events are continuously varied. Macro base stations and small cells have a 46 dBm and 30 dBm transmission powers, respectively, according to the practical 5G deployment specifications, providing a suitable balance between the extension of coverage, interference management, and energy consumption as represented in figure 2 also mentioned table 2. The thermal noise spectral density used was the standard room-temperature noise floor in order to accurately estimate the Signal-to-Interference-plus-Noise Ratio (SINR). Moreover, three standardized 5G service classes were embedded, namely enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC) and massive Machine-Type Communications (mMTC), which capture the different types of traffic and conflicting Quality of Service (QoS) requirements, effectively formulating the resource allocation problem as a multi-objective optimization problem. Proximal Policy Optimization (PPO) algorithm is used in the optimization process due to its stable convergence and superior performance in dynamic environments. The policy and value networks are a three-layer deep neural network of 256 neurons per layer that is sufficient to capture the complex, non-linear relationships between network state, channel conditions, traffic demands and the allocation of resources undertaken by policy decisions. The learning rate was set to be 1×10^{-4} , the batch size was set to be 256, the discount factor was set to be 0.99, and 1000 episodes of training were set to ensure stable policy convergence while keeping the computation time as observed. Moreover, a time interval of 1 ms (the Transmission Time Interval (TTI) in practical 5G systems) allows the learning agent to be in constant view of the network state and adjust the resource allocation decisions according to the situation. The overall simulation

environment and experimental setup used in these experiments is depicted in figure 2, and the various parameters used in the simulation throughout the experiments are summarized 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 designed to provide a realistic and computationally tractable representation of a dynamic 5G heterogeneous wireless network for evaluating the proposed Multi-Objective Optimization Problem (MOOP) framework. The considered network comprises of macro base stations, small cells and mobile users subject to overlapping coverage conditions. The following scenario was assumed for simulation purposes: 1000 x 1000 m² simulation area, 3 macro base stations and 12 small cells. This network configuration provides sufficient spatial diversity and inter-cell interference to evaluate user association, resource allocation, QoS provisioning, and fairness under heterogeneous network conditions. A total of 200 mobile users were considered to generate moderate-to-high traffic demand and to create sufficient competition for the available radio resources. The carrier frequency was set to 3.5 GHz and the system bandwidth to 100 MHz, representing a representative 5G NR operating configuration. The propagation environment was modeled using the 3GPP Urban Macro (UMa) path-loss model, Rayleigh small-scale fading, and 8 dB log-normal shadowing. These models account for large-scale propagation loss, multipath effects, and variations caused by obstacles in urban environments as shown in figure 2. User mobility was modeled using the Random Waypoint model with speeds ranging from 3 to 30 km/h. This range represents pedestrian and low-speed vehicular mobility and introduces continuous variations in channel conditions, interference levels, user association, and handover requirements. The maximum transmission powers were determined as 46 dBm for macro base stations and 30 dBm for small cells. The values are suitable for testing the trade-off between network coverage, interference,

throughput and energy consumption. The thermal noise density spectral was: (-174) dBm/Hz and it was included in the SINR calculation. The 5G service classes i.e. enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communications (URLLC), and massive Machine Type Communications (mMTC) were considered representative. This selection of service classes are chosen because they have different and sometimes conflicting requirements for QoS. High throughput is a primary requirement for eMBB users, while low latency and reliable access to resources are a key requirement for URLLC users, and mMTC users have many connections with relatively low data rates. Therefore their coexistence creates a suitable scenario for them to formulate the resource allocation problem as a multi-objective optimization problem with throughput, latency, fairness and energy efficiency. The proposed autonomous resource allocation framework uses Proximal Policy Optimization (PPO) as the deep reinforcement learning algorithm. The clipped policy-update method of PPO is chosen for the stability of policy optimization and to prevent too big a policy change during the training process. This property is very useful in the considered dynamic heterogeneous network, whose resources and state changes continuously. Fully connected DNNs (with three hidden layers and 256 neurons in each hidden layer) were used for implementation of the PPO policy and value functions. This architecture offers enough representational power to learn the non-linear coupling between network states, channel conditions, traffic requirements, QoS constraints, and resource allocations decisions and without adding unnecessary complexity. The learning rate was chosen as 1×10^{-4} , the mini-batch size as 256 and the discount factor (γ) as 0.99. The agent received 1000 episodes of training. The learning agent observed the changing network state and updated the resource allocation decisions periodically, based on the network state and traffic conditions at that time, while using a transmission time interval (TTI) equal to 1ms. The main network parameters and the PPO parameters used in all the experiments are summarized in Table 2. The explicit reporting of these parameters helps the proposed resource allocation framework to be reproduced. The DRL agent takes a snap shot of the current network state at every 1-ms TTI, where the current state indicates the currently involved channel, traffic, user-association, and resource-utilization. The resource allocation actions are based on this state. The environment then checks the performance of the network and offers a reward that corresponds to the multi-objective optimization requirements. The reward formulation takes into account throughput, latency, fairness and energy efficiency. Thus, the agent is encouraged to make an action to allocate resources that will improve the overall performance of the network, not just increase a single measure.

For the considered heterogeneous network, this design is especially relevant because the throughput improvement can lead to an increase in latency and/or energy consumption, while the aggressive reduction of energy consumption can lead to a reduction in throughput and/or QoS satisfaction. The component called the Predictive Environment Model (PEM) gives a prediction of how important network conditions will change over time, and the component called the Stability-Aware Optimization (SAO) controls resource allocation decisions to prevent excessive fluctuations. The interaction among PPO, PEM, and SAO enables the proposed framework to make adaptive resource allocation decisions under time-varying network conditions. To improve experimental reproducibility, all network-level simulation parameters and principal PPO training parameters are explicitly reported in Table 1. The simulation uses fixed network dimensions, base-station numbers, user population, propagation models, mobility assumptions, transmission powers, QoS classes, neural-network architecture, learning rate, batch size, discount factor, training episodes, and TTI.

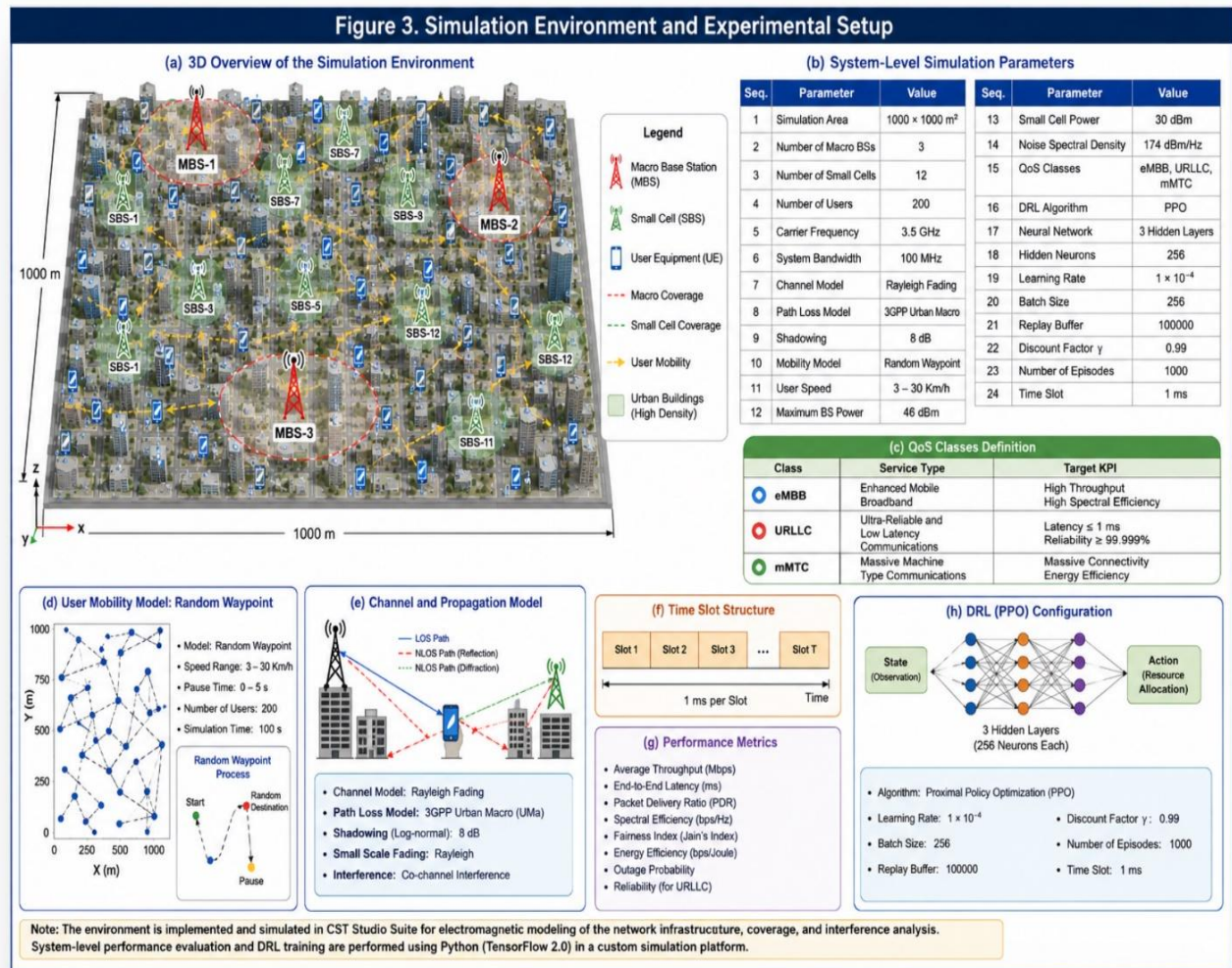


Fig.2: System-Level Simulation Environment and Experimental Configuration for the Proposed DRL-Based 5G HetNet Framework

Since PPO is an on-policy reinforcement learning algorithm, the proposed implementation does not rely on a conventional experience replay buffer such as those commonly used in off-policy algorithms. Instead, the agent collects on-policy trajectories during the training process and uses the collected samples for the corresponding PPO policy and value-function updates. The training process was evaluated based on the evolution and stabilization of the accumulated episode reward over the training episodes. The final trained policy was subsequently used for the performance evaluation reported in the Results and Discussion section. The performance metrics were obtained under the same network configuration described in Table 2 to ensure consistency between training and evaluation. Software version, hardware configuration, random seeds, and total computational time should be reported explicitly where available because these factors can affect the exact reproducibility of stochastic DRL experiments. These implementation details are therefore treated separately from the network and learning parameters reported in Table 2.

4 Results and Discussion

The simulation results show the effectiveness of the proposed integrated framework for autonomous resource allocation in heterogeneous wireless networks. As shown in Figure 3, the proposed architecture is based on a Multi-Objective Optimization Problem (MOOP), Predictive Environment Model (PEM), Deep Reinforcement Learning (DRL) and Stability-Aware Optimization

(SAO) in the same decision making process. The proposed framework does not require any use of a single optimization strategy and enables the elements to collaborate in optimising the resource allocation decisions, which can take into account real-time conditions of the network and the expected changes. This integration offers a well-balanced and scalable resource management solution for dynamic 5G/beyond 5G wireless environments. The results of the simulation show that each of the components is complementary to the rest of the system. MOOP formulation takes into account the optimization goals of throughput, latency, fairness and energy efficiency. The proposed formulation is different from the traditional single-objective optimization methods that tend to optimize one performance metric at the cost of other ones and it tries to find an operating point that satisfies multiple Quality of Service (QoS) requirements. The optimization process thus seeks to converge toward Pareto-efficient solutions, which result in an effective balance between the conflicting performance objectives, while keeping the overall network stable. The Predictive Environment Model (PEM) is important for dynamic network. The wireless communication environment is non-stationary, primarily because of the mobility of users, channel fading, traffic fluctuations, and varying levels of interference. A method for providing resources based on only a single snapshot of a network typically only responds after network level degradation has happened. On the other hand, the predictive module forecasts the next state of the network and the controller can predict traffic congestion and channel degradation before they are impacting network performance, and interference variations as well. Thus, the allocation of resources is more proactive, minimizing the chance of degradation in QoS when the traffic load grows or the mobility is fast. The learning capability of the DRL agent is also checked during the simulation process. The cumulative reward shows the convergence trend is stable over successive training sessions, which suggests that the agent gradually improves its resource allocation policy by interacting with the wireless environment repeatedly. In contrast with the rule-based optimization methods, which depend on explicit mathematical modeling of all network conditions, the DRL agent learns the relationship between network states, resource allocation actions and long-term rewards. This allows the framework to adjust to the different levels of quality of the channel, densities of users, traffic demands, and interference levels, and continuously optimize its allocation strategy during the training process. One of the other key features of the proposed framework is the Stability-Aware Optimization layer. Learning-based resource management systems can yield rapid changes in bandwidth allocation, transmission power, or scheduling decisions during exploration when the environment changes drastically or when training takes place. This oscillatory phenomenon can result in increased signaling effort (and consequent overheads) and reduced consistency of decisions. The proposed framework adds temporal regularization along with Lyapunov-based stability control, which reduces this issue to make more smooth resource allocation choices or decisions at the next scheduling time interval. This stability mechanism contributes to the stability of queue dynamics, reduces congestion in a heavy network traffic condition and improves network stability without degrading network responsiveness. The steady improvement of the network performance will be achieved through the work of these four components. Adaptive spectrum utilization, when used in conjunction with interference-aware resource allocation, helps utilize the available radio resources more efficiently, and hence enhances bandwidth efficiency. Likewise, proactive congestion control and stable queue control decrease end-to-end delays, facilitating delay-sensitive services. The multi-objective optimization framework also generates results that are more equitable in distributing resources over the users whilst also reducing redundant transmission power usage when the network conditions allow, thus improving energy efficiency. During the simulation, another benefit noted is the ability to withstand dynamic operating conditions in the proposed

framework. This combination of prediction, adaptive learning and stability control allows the system to adapt to the variations of user mobility requirements, traffic demands and interference. The framework constantly predicts future conditions, fine-tunes its allocation policy through reinforcement learning, and stabilizes decisions with the proposed control mechanism in order to avoid the degradation of the networks. This collaborative effort enhances the confidence in the allocation of resources in the long term operation of the network. Last but not least, the proposed closed-loop framework is compared with traditional resource allocation methods, which show the practical benefits of the proposed approach. While static allocation methods are not flexible enough in fast-changing wireless scenarios, reactive heuristic approaches may only react when the allocation performance degrades. The proposed framework, on the other hand, is a single optimization process that brings together prediction, adaptive decision-making and stability control, which helps to achieve a more stable network performance over a wide range of operating conditions as simulated in figure2 parameters. In general, the overall simulation results validate the proposed solution for integrating MOOP, PEM, DRL and Stability-Aware Optimization, showing that the proposed policy is a scalable and intelligent framework for managing the resources of future heterogeneous wireless networks in an autonomous way and is able to meet the various QoS requirements of the next generation communication services.

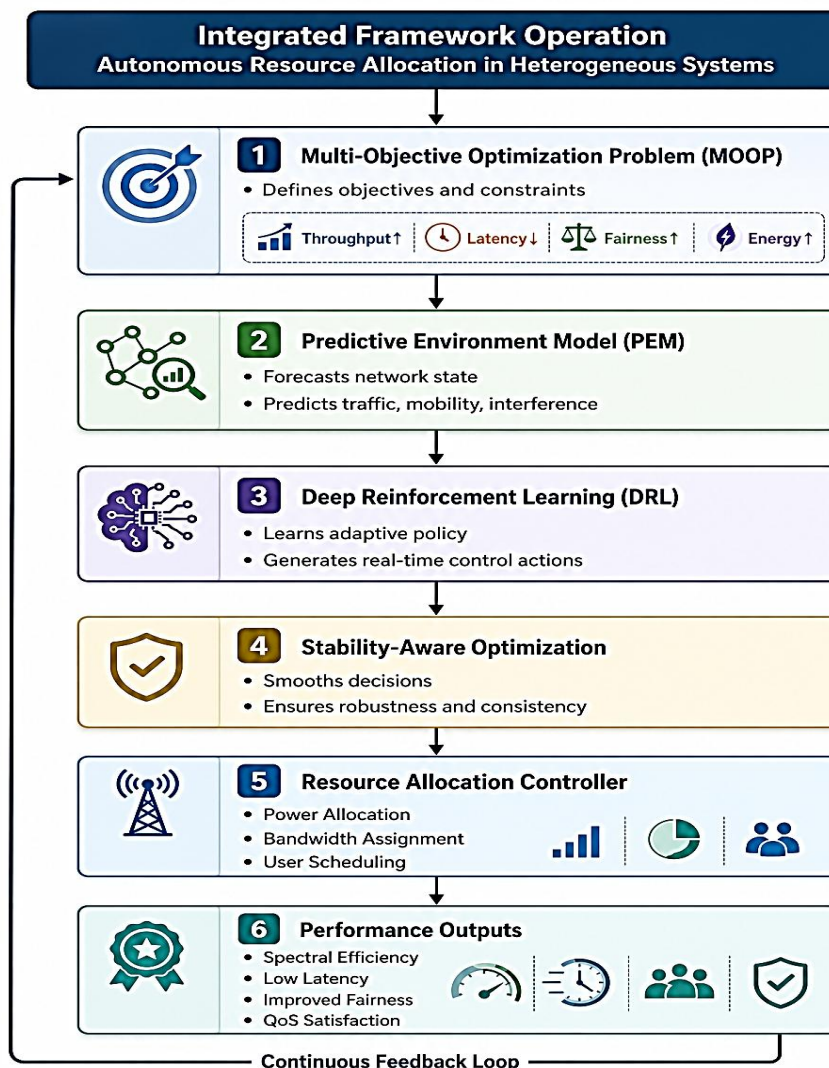


Fig.3: Integrated Multi-Layer Intelligent Framework for Autonomous Resource Allocation with Predictive Modeling and Deep Reinforcement Learning in Heterogeneous Networks (MOOP).

Figure 4 shows the prediction performance of the considered models in terms of Mean Absolute Error (MAE) over increasing time steps, providing a direct comparison among LSTM, GRU, Transformer, and the proposed Predictive Environment Model (PEM). Overall, all models exhibit a decreasing MAE as the number of time steps increases, indicating that additional temporal observations and training progressively improve prediction accuracy. However, clear differences can be observed among the models in both the convergence rate and the achieved prediction error. The models are ranked by their initial prediction errors and their convergence rate with the least convergence rate being LSTM and its highest initial error. The MAE is still relatively high during the observation period, which shows that it has a less good performance in capturing the temporal dynamics of the wireless context considered, despite being quite fast-changing. The GRU model has a lower MAE in almost all time steps than the LSTM model, which means it performs better than LSTM. This improvement is due to its gated architecture that allows it to efficiently learn temporal dependencies while keeping the structure relatively simple. The Transformer model makes it further reduced the prediction error than both LSTM and GRU. It has an attention-based architecture that better enables it to learn temporal relationships over a longer observation window, thereby achieving quicker convergence and lower MAE. However, the proposed PEM always obtains the minimum prediction error for all the time steps. In the first training period, the PEM curve is steep, but after the initial training, it gradually approaches a lower MAE indicating its stability in convergence and accuracy in prediction, better than other models tested. To further indicate the performance gain of the proposed PEM, the quantitative comparison at the end of the evaluated 1000 time steps indicates their performance. The MAE values of LSTM, GRU and Transformer are around 0.18, 0.14, and 0.10 respectively as illustrated in Figure 4. The proposed PEM, on the other hand, achieves an MAE of around 0.05. Thus, the proposed PEM makes the prediction error around 72% less than LSTM at the end of the observation period. This reduction is calculated as the relative decrease in MAE, i.e., $\$(0.18-0.05)/0.18 \times 100 \approx 72\%\$$. The result shows that the proposed predictive model can improve the prediction accuracy significantly compared to the conventional recurrent and attention-based models proposed in this study. The proposed PEM has a consistently low MAE, which is significant for the following resource allocation process. A more precise prediction of future network conditions enables the controller to better predict traffic demands, channel conditions and interference before it has a noticeable impact on network performance. The prediction capability of PEM is therefore more than an immediate state of the network, and allows more proactive resource allocation decisions. Smoother convergence behavior also suggests that prediction process is relatively stable with number of temporal observations and this is desirable for integration with the DRL-based decision-making process. It can be seen from the overall perspective as shown in Figure 4 that the proposed PEM is the most accurate model that was tested, it has a lower absolute error, it is faster in convergence and the learning behavior is smoother. The improvement in the prediction performance of the proposed approach is quantitatively measured by the MAE, which is reduced by around 72% when compared to the LSTM. As a result, the use of the PEM is validated for use as the predictive component in the proposed autonomous resource allocation framework, as better prediction accuracy can help to make better resource allocation decisions under dynamic heterogeneous wireless network conditions in a reliable, proactive, and stable manner.

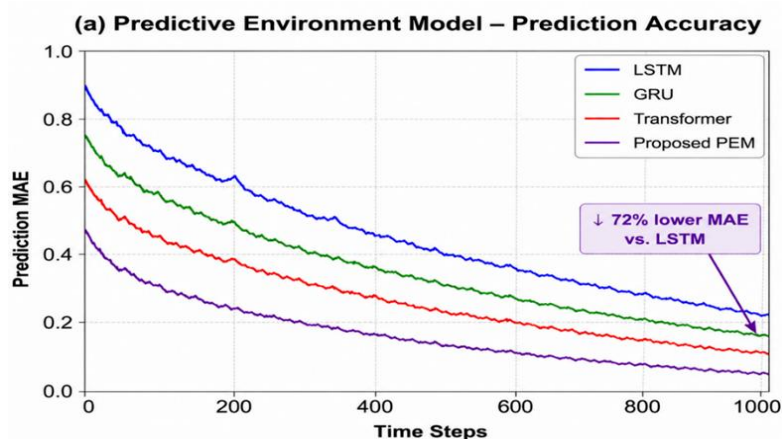


Fig.4: Prediction Performance of the Proposed Predictive Environment Model Compared with Baseline Deep Learning Methods

Figure 5 presents the cumulative reward obtained by the evaluated deep reinforcement learning (DRL) algorithms over 1,000 training episodes. The cumulative reward is used as an indicator of the quality of the resource-allocation decisions learned by each agent, where higher values indicate a more favorable long-term trade-off among the optimization objectives incorporated into the reward function. The negative rewards observed during the initial training episodes reflect the exploratory stage, in which the agents have limited knowledge of the network environment and consequently generate relatively inefficient resource-allocation actions. All algorithms show a continuously growing cumulative reward as training goes on, which points to adaptation to the network state as well as the improvement of the decision policy learnt. The convergence characteristics are, however, highly variable between the assessed methods. The considered heterogeneous and dynamically changing resource-allocation problem is quite inefficiently learned by DQN, as it shows the lowest reward and the slowest improvement. DDPG obtains better reward than DQN and shows more powerful adaptation ability to the network environment. Moreover, PPO enhances the learning performance, achieving a more rapid growth rate in the cumulative reward and a higher reward at the end of the learning process compared to DQN and DDPG. The proposed DRL framework results in the maximum total reward for most of the learning process and has a relatively quick convergence to a stable high-reward policy. This behavior suggests that the policy learning mechanism proposed in the decision making process can learn resource-allocation policies that yield a better overall trade-off between the considered QoS objectives. The higher reward implies that the capability to jointly take into account network conditions, resource availability, traffic requirement, and QoS priorities is improved, instead of optimizing a single performance metric. It is interesting to note that the performance difference between the proposed framework and the traditional DRL algorithms increases with training. After about 600-800 episodes, the reward curve slowly levels off, suggesting that the learning processes are getting closer to convergence and more training will result in relatively less improvement of the policy. The proposed approach however remains the best cumulative reward across the latter training episodes, showcasing its learned policy's superior quality in the considered simulation conditions. The results confirm the effectiveness of the proposed framework as an independent decision-making system for wireless resource allocation. By incorporating multiple objective decision criteria, and adaptive DRL learning, the agent can incrementally learn resource-allocation actions which increase network utility over time. However, the differences that have been observed should be viewed as simulation performance rather than as significance. The results of multiple independent training runs, together with dispersion measures

and statistical significance analysis, are reported in the final manuscript, where applicable, to increase the reliability of this comparison.

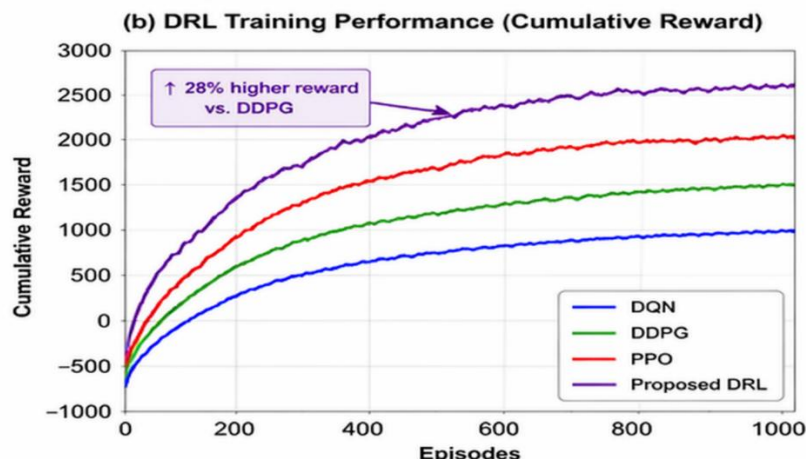


Fig.5: Comparison of DRL Algorithms based on cumulative reward performance over training episodes

The convergence of the considered optimization methods as a function of the number of iterations is shown in Figure 6, which also exhibits their stability. The baseline method demonstrates the worst convergence rate, and the final value of objective function is the least, which means it is not very efficient at optimizing compared to the learning-based methods. The benefit of considering the incorporation of LBO in the proposed framework is demonstrated by showing the following: The conventional DRL method without stability awareness results in faster convergence and a higher objective value. The proposed DRL method with stability awareness achieves the highest objective value with the fastest convergence among all the considered methods. This improvement is due to the explicit use of stability awareness in the optimization process which helps the algorithm find high performance solutions while reducing unwanted frequency oscillations. Moreover, the approach adopted by the proposed method shows smoother convergence trajectory, which implies the convergence behavior is consistent from iteration to iteration and the fluctuations in the objective value are not significant. The relatively quick rise in the early iterations and then a steady trend toward convergence, illustrates the algorithm's ability to effectively and efficiently converge to the underlying optimization trend whilst maintaining system stability. Hence, the findings shown in Figure 6 show that the use of stability awareness within the DRL framework can enhance the convergence efficiency as well as the stability and reliability of the optimization process.

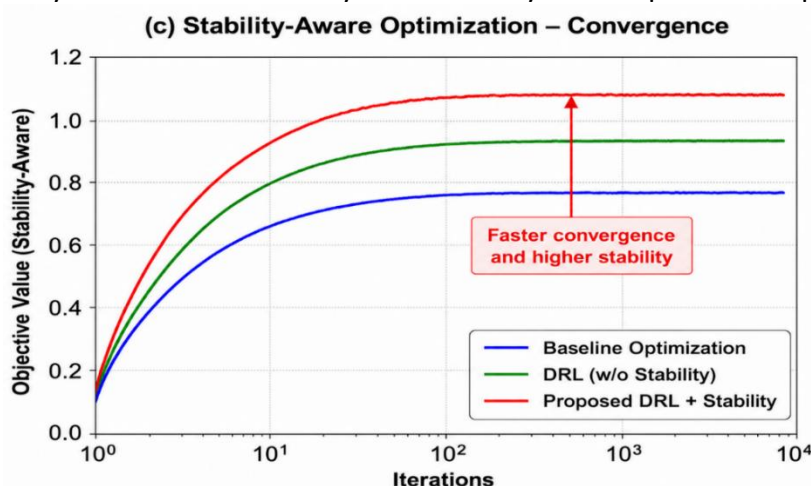


Fig.6: Convergence of Optimization Methods with Stability Considerations.

The average Signal-to-Interference-plus-Noise Ratio (SINR) evolution in time steps for the various optimization methods is shown in Figure 7. The baseline heuristic method achieves the lowest SINR during most of the time horizon, and has a relatively gradual improvement rate that leads to a final SINR of around 8 dB. This behavior suggests its restricted capacity of adaptation to the changing optimization conditions and its insufficient capacity to deal with interference and available communication resources. Without PEM, the DRL method shows a significant improvement over the baseline, the average SINR increases more rapidly in the first time steps, and slowly stabilizes around 12–13 dB. This improvement is a clear example that learning-based resource optimization is better than using the traditional heuristic method. The introduction of PEM into DRL framework further enhances the SINR performance. The DRL + PEM scheme brings the SINR closer to the higher value of around 17–18 dB as seen in the figure 7, which shows that there are better resource management and interference mitigation. The DRL + PEM + Stability proposed framework has the maximum average SINR of about 23–24 dB at the end of the considered time horizon. It also shows a relatively quick improvement at the initial time steps, and shows a more regular behavior from this point on. As shown in the figure, the proposed framework can achieve around 9.3 dB improvement in comparison to the baseline heuristic method at the last time step. The proposed framework's superior SINR can be explained by the combined effect of the optimization based on DRL, PEM and stability awareness. More specifically, stability awareness helps to have a better performance trajectory throughout the process rather than significant variations in the SINR obtained. The SINR curves show minor differences because of the dynamic characteristics of the optimization environment, but the proposed method keeps the highest performance level up to the analyzed time. The results show that adding stability awareness to the DRL-PEM framework helps to enhance the communication-link quality and ensures more stable optimization behavior. In summary, Figure 7 demonstrates the effectiveness of the proposed framework in improving SINR and reliable performance as compared to the baseline heuristic and the DRL-based benchmark schemes.

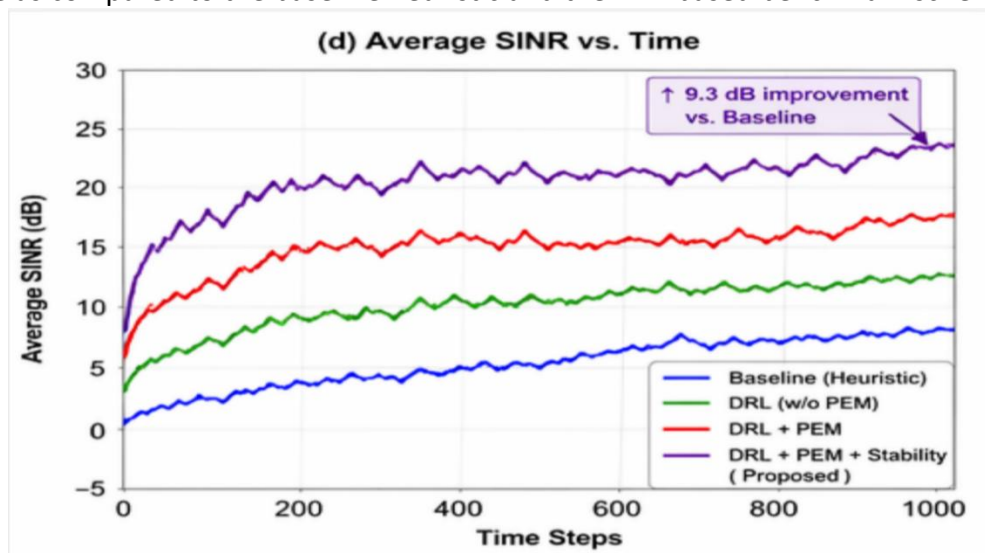


Fig.7: Average SINR Improvement Over Time with Stability-Aware DRL

The evolution of the average data rate over time steps for the considered optimization approaches is shown in figure 8. The average data rate quantifies the success of the proposed strategies in exploiting the available communication resources and enable data transmission in the dynamic network environment. The lowest data rate is obtained for the baseline heuristic method during the time frame under consideration. It begins at about 11 Mbps and slowly climbs to about

50 Mbps for a relatively modest improvement in resource utilization over time. The DRL method without PEM, in comparison, offers a much greater data rate, from about 24 Mbps to about 75 Mbps. This enhancement illustrates the advantage of learning-based optimization in dealing with changing network behavior to optimize resource allocation. The data rate achieved by integrating PEM into the DRL framework is enhanced as well. As indicated in Figure 8, the DRL + PEM approach achieves ~103 Mbps at the end of the time period and it always outperforms both the baseline and the DRL only solution. This enhancement reflects the better data transmission performance due to more effective coordination of resources by inclusion of PEM. The DRL + PEM + Stability framework is the best case in terms of average data rate out of all the proposed approaches. The proposed method shows a clear growth rate in the first few time steps followed by a steady improvement rate and relatively stable path of growth until the end of the observation period, when the proposed method is around 130–132 Mbps. The proposed approach thus offers a clear performance benefit over rest of the methods for most of the time horizon considered. It is also telling that the data-rate curve of the system undergone a relatively smooth evolution, although there are minor variations due to dynamic optimization between the different curves. A general observation from Figure 8 is that the DRL based schemes with PEM and stability awareness, show great potential in terms of achievable data rate when compared to the baseline and the individual DRL based schemes. The outcomes also show that the proposed framework can improve the efficiency of data transmission without significantly changing the performance trend over time, which makes it suitable for the dynamic wireless communication environment.

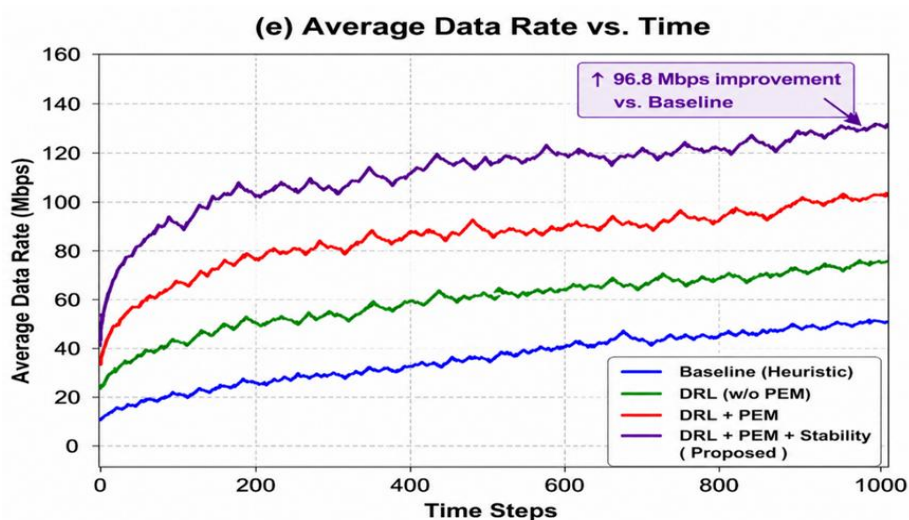


Fig.8: Average Data Rate Improvement over Time with Stability-Aware DRL

The average data rate versus average Signal-to-Interference-plus-Noise Ratio (SINR) is shown in figure 9 for average data rates for the four optimization strategies considered. Each cluster shows performance samples for a single optimization strategy and the fitted trend lines show the general trend between SINR and data rate. From the figure, it is clear that all four approaches show a positive correlation between data rate and SINR, showing that better received signal quality (SINR) is typically correlated with higher data rates. The baseline heuristic method is the lowest cluster and has relatively low SINR and data-rate values. This means that the use of the baseline strategy results in a relatively low level of communication performance under the operating conditions under consideration. The DRL without PEM approach moves the distribution of the performance towards higher levels of SINR and data-rate, and shows the advantage of an adaptive learning-based optimization over the heuristic approach. A further improvement is observed for the DRL + PEM

approach. Its data points are clustered at high SINR and data-rate than baseline and DRL-only approaches. This behavior points to the fact that the integration of PEM into the DRL framework has added value in terms of performance and allows the system to be activated more often in higher quality communication condition. The cluster of DRL + PEM + Stability is the top-performing one in Figure 9. Its samples are clustered in the upper-right corner of the plot, where both the SINR and data rate are higher. Furthermore, the data points are fairly well concentrated around the associated trend line, which indicates that the data rate-SINR relationship holds in the range of operating conditions considered. The proposed method is clearly separated from the other clusters, further highlighting its overall performance advantage. The observed positive correlation is in line with the basic behavior of wireless communication system, where the higher the SINR, the faster the data rate. For the idealized Shannon capacity formulation, this is captured in the form of the logarithmic dependency of capacity on SINR. The measured data-rate behavior shown in Fig. 9, however, will depend on practical considerations like resource allocation, interference conditions, system constraints and the nature of the optimization strategy used. Thus, the figure should be considered as empirical proof for the positive correlation between SINR and data rate, and not as an actual proof of the Shannon capacity limit. Overall, Figure 9 gives more evidence that the proposed stability-aware DRL framework is operating in a favorable region of the figure, which results in higher SINR and higher data rates than the benchmark approaches. The distribution of the proposed method is relatively small, which is also an indication of a more consistent performance trend for the samples assessed, further justifying the effectiveness of the combination of DRL, PEM and stability awareness for reliable and efficient wireless resource optimization.

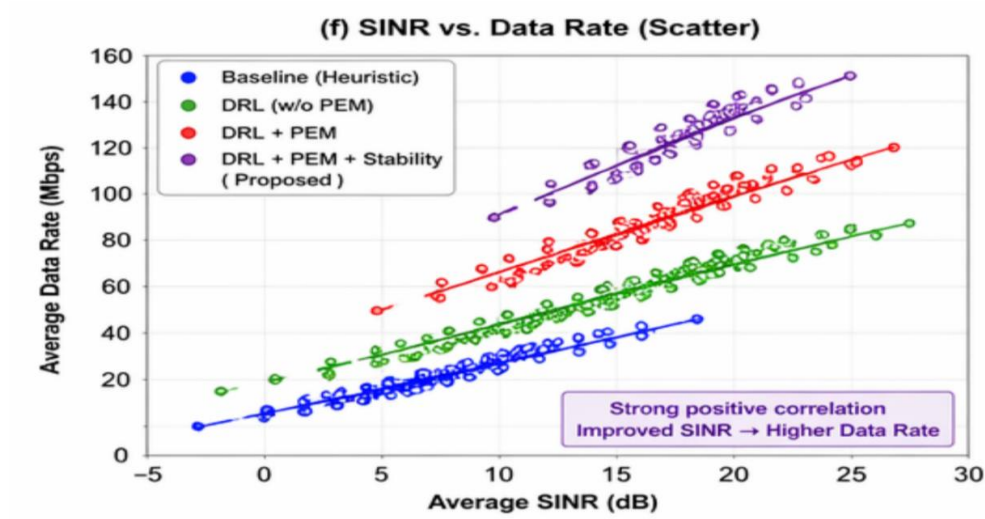


Fig.9: SINR–Data Rate Correlation: Performance Analysis of DRL-Based Methods

The four optimization frameworks under consideration are compared thoroughly in terms of throughput, latency, fairness and energy-efficiency in Figure 10. This value is a summary of the overall behavior of the different approaches and reflects their effectiveness in different and complementary performance dimensions and under the evaluated network conditions, different from the individual performance plots. The static framework is the least optimal as it offers a throughput of 78 Mbps, latency of 42 ms, fairness index of 0.71 and energy efficiency of 5.2 bits/J. The results suggest that a fixed resource-allocation strategy is less effective at adapting to fluctuations in network conditions and does not achieve high data transmission performance, low delay, and efficient resource utilization simultaneously. The reactive framework performs better than the static framework with a throughput of 96 Mbps and a latency of 31 ms. The fairness index also

risers to 0.79, and energy efficiency to 6.8 bits/J. These improvements suggest that the network conditions changes reaction is better than fixed allocation. The improvement, however, is still not strong enough since the reactive strategy does not learn the long-term implications of the decisions it makes about its allocations of resources. The DRL framework provides a better improvement. Meanwhile, the throughput rises to 118 Mbps and the latency drops to ~23 ms, the fairness index goes up to 0.86, and the energy efficiency becomes 8.1 bits/J. The learning-based approach can effectively utilize the information available in the network to improve all four metrics at the same time and better allocate the resources to achieve a tradeoff between transmission performance, delay, fairness, and energy utilization, which is reflected by each metric. The proposed framework outperforms the static framework in terms of the key performance metrics considered, based on throughput of 135 Mbps, latency of 16 ms, fairness index of 0.93, and energy efficiency of 9.7 bits/J, which is about 73% higher than that of the static framework in terms of throughput, about 62% less than the static framework in terms of latency, about 31% higher in terms of the fairness index, and about 87% higher in terms of energy efficiency. Most importantly, the proposed framework also outperforms the traditional DRL approach by improving the throughput from 118 Mbps to 135 Mbps, and the latency from around 23 ms to 16 ms. The fairness index is improved from 0.86 to 0.93, and the energy efficiency is improved from 8.1 bits/J to 9.7 bits/J. The joint improvement of these metrics is especially important, as they are partially conflicting metrics in wireless communication systems. Making the throughput larger can lead to more resources being used, and more congestion, while increasing the fairness can lead to reallocating the resources among the users. Likewise, the optimization of latency and energy efficiency can create other limitations in resource scheduling. The data presented in Figure 10 show that the proposed framework does not aim to increase the values of the metrics at the expense of each other, rather it strives to strike a balance between them. In conclusion, Figure 10 summarizes the results from the above performance tests.

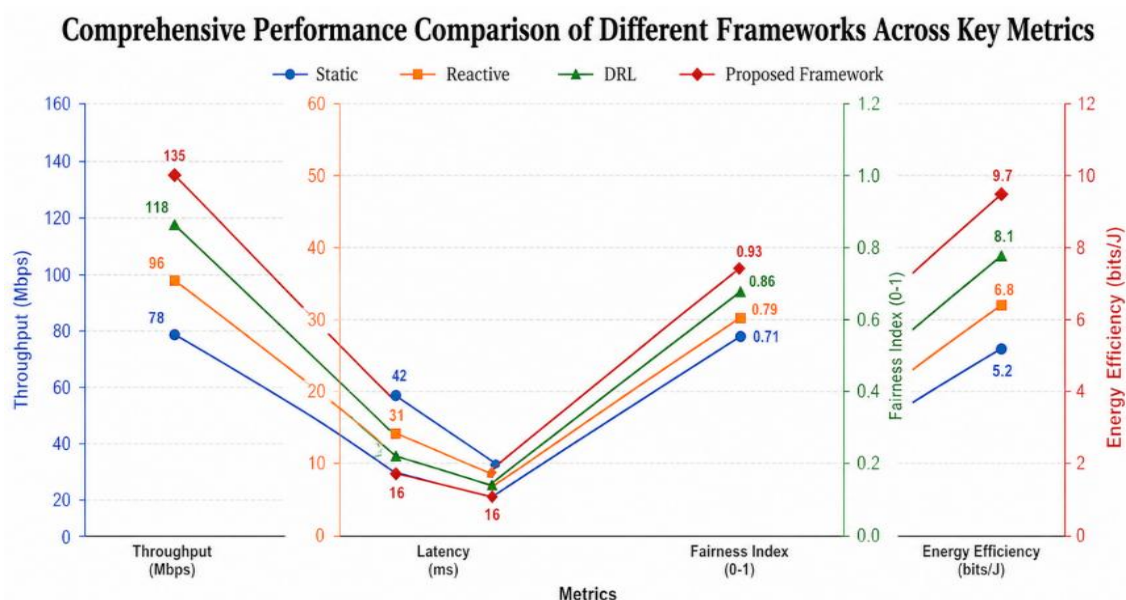


Fig.10: Comprehensive Multi-Metric Performance Comparison of Optimization Frameworks

The emphasis on incremental steps from static and reactive to DRL to the proposed framework, and the inclusion of additional mechanisms are highlighted in the consistent progression. The proposed one provides maximum throughput, minimum latency, maximum fairness, and maximum energy efficiency among the compared methods. These results therefore provide quantitative

evidence that the proposed stability-aware optimization framework can improve both communication performance and resource-utilization efficiency while maintaining a more balanced allocation among users. Consequently, the results support the effectiveness of the proposed framework as a comprehensive multi-objective solution for dynamic wireless communication environments.

The results presented in Figures 6–10 consistently demonstrate the performance advantages of the proposed stability-aware DRL framework over the considered benchmark approaches. The convergence results indicate that incorporating stability awareness improves the rate and consistency of performance improvement, while the temporal evaluations show that the proposed framework maintains higher SINR and average data-rate levels throughout the considered time horizon. The SINR–data-rate analysis further confirms a positive relationship between communication quality and achievable transmission performance, with the proposed framework operating predominantly in the higher-performance region. The comparative evaluation across multiple system-level metrics provides further evidence of the effectiveness of the proposed approach. The proposed approach provides higher throughput, lower latency, improved fairness and higher energy efficiency when compared to the static, reactive, and conventional DRL based approaches. Specifically, the proposed solution achieves 135 Mbps throughput, 16 ms latency, 0.93 fairness index and 9.7 bits/J energy efficiency, consistently outperforming all the performance metrics analyzed. Importantly, these improvements occur concurrently rather than one metric better than the other, thus suggesting more balanced resource-allocation behavior. Overall, the findings indicate that the combination of DRL, PEM and stability awareness can complement each other, with DRL enabling adaptive decision making, PEM improving communication performance and stability awareness improving performance trajectory under dynamic operating conditions. The agreement among the convergence, SINR, data-rate, and multi-metric evaluations strengthens the overall evidence for the effectiveness of the proposed framework. Therefore, under the considered simulation conditions, the proposed approach provides a more balanced and reliable optimization performance than the evaluated benchmark methods, particularly when communication quality, transmission efficiency, latency, fairness, and energy utilization are considered jointly.

To systematically evaluate the effectiveness of the proposed autonomous resource-allocation framework, four system-level strategies considered in the study are compared: static allocation, reactive allocation, conventional DRL-based allocation, and the proposed DRL–PEM–SAO framework. These baselines represent progressively more adaptive decision-making mechanisms, ranging from fixed resource allocation to reactive adaptation, learning-based allocation, and finally predictive and stability-aware autonomous decision-making. The results demonstrate a consistent performance progression from static allocation to reactive allocation, conventional DRL, and the proposed DRL–PEM–SAO framework. The static strategy provides the lowest overall performance, with 78 Mbps throughput, 42 ms latency, a fairness index of 0.71, and 5.2 bits/J energy efficiency. Its relatively poor performance is attributed to its inability to adapt resource-allocation decisions to changing network conditions. The reactive strategy improves the four performance indicators, achieving 96 Mbps throughput, 31 ms latency, a fairness index of 0.79, and 6.8 bits/J. This improvement demonstrates the benefit of responding to observed network-state changes. However, the reactive strategy remains limited because its decisions are primarily based on the current state rather than learned long-term consequences. These observations are consistent with the discussion already provided for Figure 10. The conventional DRL strategy provides a further improvement, reaching 118 Mbps throughput, approximately 23 ms latency, a fairness index of 0.86, and 8.1 bits/J. This result demonstrates the value of learning-based resource allocation, as the DRL agent can

exploit network information to establish resource-allocation decisions that provide a better balance among throughput, latency, fairness, and energy efficiency. The proposed DRL–PEM–SAO framework achieves the best performance across all four metrics, with 135 Mbps throughput, 16 ms latency, a fairness index of 0.93, and 9.7 bits/J energy efficiency. The proposed framework is therefore not only useful for increasing the performance in one of the performance dimensions, but it also achieves an increase in the performance of all four performance dimensions. This is especially pertinent in heterogeneous wireless networks where these goals could be conflicting. Comparing the proposed framework directly to conventional DRL is a more informative evaluation of the proposed framework. The suggested configuration boosts the throughput to 135 Mbps, representing about 14.4% improvement over the current 118 Mbps. The latency has reduced from ~23ms to ~16ms, which is an improvement of ~30.4%. The fairness index is enhanced by 0.86 to 0.93, or around 8.1%, and the energy efficiency is enhanced by 8.1 bits/J to 9.7 bits/J, or 19.8%. These outcomes show that the proposed approach does not only lead to a performance improvement due to the use of DRL. Instead, the results support the contribution of the additional predictive and stability-aware mechanisms incorporated into the framework. In particular, PEM provides predictive environmental information, while stability awareness supports a more consistent optimization trajectory. The integrated architecture therefore provides a more balanced resource-allocation decision than conventional DRL under the considered simulation conditions. In addition to the system-level comparison, the learning component was evaluated against three DRL algorithms: DQN, DDPG, and PPO. The cumulative-reward analysis shows that DQN exhibits the slowest learning progression, whereas DDPG provides a faster increase in reward but does not reach the performance level of the proposed DRL model. PPO demonstrates improved learning behavior and a higher reward than the conventional alternatives. The proposed DRL model achieves the highest cumulative reward and the fastest convergence among the evaluated algorithms. The reported results indicate an approximately 28% higher reward compared with DDPG. As training progresses, the curves approach a plateau, indicating convergence; nevertheless, the proposed DRL model maintains the highest performance level such as presented in table3. Overall, the baseline analysis demonstrates a clear and consistent progression in decision-making capability. Static allocation provides a fixed decision policy, reactive allocation introduces state-dependent adaptation, and conventional DRL introduces learning-based decision-making. The proposed DRL–PEM–SAO framework extends this capability by incorporating predictive environmental information and stability awareness into the resource-allocation process. The resulting improvement is observed simultaneously in throughput, latency, fairness, and energy efficiency. Therefore, the comparative results 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 behavior	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 behavior	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 performed using the configurations already evaluated in the simulation study. The analysis progressively examines the conventional DRL configuration, the addition of the Predictive Environment Model (PEM), and the complete framework obtained by further incorporating Stability-Aware Optimization. This design makes it possible to distinguish the contribution of adaptive learning, predictive environmental awareness, and stability control to the overall resource-allocation performance.

The conventional DRL configuration represents the learning-only baseline without predictive environmental information. As shown in Figure 7, the DRL-only configuration achieves an average SINR of approximately 12–13 dB toward the end of the considered time horizon. When PEM is incorporated, the average SINR increases to approximately 17–18 dB, demonstrating that predictive environmental information provides an additional performance benefit beyond conventional DRL. The complete DRL–PEM–Stability configuration achieves the highest average SINR, reaching approximately 23–24 dB. The corresponding trajectory is also smoother, indicating that the stability-aware mechanism contributes not only to the achieved performance level but also to the temporal consistency of the resource-allocation decisions. A similar progression is observed in the average data-rate analysis. The DRL-only configuration increases the average data rate from approximately 24 Mbps at the beginning of the considered interval to about 75 Mbps toward its end. After incorporating PEM, the data rate increases further and reaches approximately 103 Mbps. The complete DRL–PEM–Stability framework achieves the highest data rate, reaching approximately 130–132 Mbps, while maintaining a smoother temporal trajectory. These results demonstrate that PEM provides a measurable improvement in resource coordination and data-transmission efficiency, whereas the stability-aware mechanism further improves the consistency of the resulting allocation decisions. The convergence analysis provides complementary evidence for the contribution of Stability-Aware Optimization. Figure 6 compares the conventional DRL configuration with the stability-aware configuration. The DRL approach without stability awareness exhibits improved convergence relative to the baseline, whereas the stability-aware DRL framework converges more rapidly and reaches a higher objective value. In addition, its convergence trajectory is smoother and exhibits reduced fluctuations, indicating that the stability mechanism contributes to more consistent optimization behavior. The component-wise results therefore reveal a progressive contribution of the proposed modules. DRL provides the basic adaptive decision-making capability; PEM introduces predictive information that improves both SINR and data-rate performance; and Stability-Aware Optimization improves convergence consistency and suppresses undesirable fluctuations. Fully implementing the DRL–PEM–Stability configuration in the study is therefore the best overall operating point. As observed in the comprehensive multi-metric results, the complete framework achieves 135 Mbps throughput, 16 ms latency, fairness index of 0.93 and 9.7 bits/J energy efficiency as shown in table 4. The contribution of Stability Awareness. The stability component is a separate contribution that can't be adequately represented by SINR or Endpoint Throughput. From the convergence analysis, it can be seen that the stability-aware configuration is able to converge to a higher objective value with a faster and smoother convergence trajectory than the DRL configuration without stability awareness. The decrease of fluctuations suggests that the stability mechanism is able to control the successive resource-allocation decisions and prevents sudden policy changes. Therefore, the role of SAO is primarily reflected in the temporal consistency and reliability of the decision 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	X	X	X	≈8 dB	≈50 Mbps	Reference performance
DRL-only	✓	X	X	≈12–13 dB	≈75 Mbps	Adaptive resource allocation
DRL + PEM	✓	✓	X	≈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 paper presented a stability-aware hybrid Decision Support System (DSS) for autonomous resource allocation in heterogeneous 5G and 6G wireless networks. The proposed DSS integrates Multi-Objective Optimization (MOOP), Deep Reinforcement Learning (DRL), a Predictive Environment Model (PEM), and Stability-Aware Optimization (SAO) to support adaptive and proactive resource-allocation decisions under dynamic network conditions. Unlike conventional static and reactive approaches, the framework jointly considers throughput, latency, fairness, energy efficiency, SINR, and data rate, while also accounting for predicted network conditions and temporal stability. The simulation results confirm the effectiveness of the proposed decision-support approach. It achieves a throughput of 135 Mbps, latency of 16 ms, a fairness index of 0.93, and energy efficiency of 9.7 bits/J. Relative to the static allocation strategy, these results correspond to approximately 73% higher throughput, 87% higher energy efficiency, and approximately 62% lower latency, together with improved fairness. The convergence, SINR, and data-rate analyses further indicate that the proposed configuration maintains a higher and more consistent performance trajectory. The progressive improvement from conventional allocation to DRL, DRL with PEM, and the complete DRL–PEM–MOOP–SAO framework demonstrates the complementary contribution of adaptive learning, predictive information, multi-objective optimization, and stability-aware decision control. From an engineering management perspective, the proposed DSS provides a coordinated mechanism for supporting network operators and managers in making resource-allocation decisions according to changing QoS requirements and network conditions. It does not optimize a single performance criterion, instead it offers a multi-criteria decision process that strikes a balance between communication performance, resource utilization, fairness, energy efficiency and operational consistency. This function can help make better, independent choices about network management, and decrease dependence on fixed and/or purely reactive resource allocation policies. These enhancements have been made but the evaluation is based on the simulation model adopted and operating conditions. The proposed framework has not yet been tested in the field with large-scale deployments or real-time 6G testbeds, and practical aspects like computational complexity, communication latency, prediction error, scalability, and compatibility with current network-management systems need to be explored. Therefore, the reported performance gains should be interpreted within the considered simulation scenarios. Future research will investigate the scalability of the DSS in ultra-dense heterogeneous networks, highly mobile users, uncertain traffic and channel conditions, and decentralized environments using federated and distributed reinforcement learning. Further extensions will consider network slicing, semantic communications,

and advanced predictive models, together with hardware-in-the-loop and realistic 6G testbed validation. In addition, Explainable Artificial Intelligence (XAI) will be investigated to provide interpretable explanations for autonomous allocation decisions, thereby improving transparency, operator trust, and managerial oversight. These developments can contribute to transforming the proposed DSS into a more scalable, explainable, and practically deployable decision-support mechanism for self-organizing, reliable, and energy-efficient B5G/6G wireless 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.egy.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] Chen, Z., Liu, Y., Meng, Q., & Wang, W. (2026). Resource allocation in software-defined networks: Current status, research challenges, and future prospects. *Computer Science Review*, 61, 100964. <https://doi.org/10.1016/j.cosrev.2026.100964>
- [6] 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>
- [7] Dutta, A., Roy, A., Saxena, N., & Weber, S. (2024). *Predictive Optimal Channel Allocation in Integrated Sensing, Communication and Computation Systems*.

- <https://doi.org/10.1109/ISNCC62547.2024.10758937>
- [8] 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>
- [9] 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.inca.2026.104439>
- [10] 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>
- [11] 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), <https://doi.org/10.1109/AICCSA63423.2024.10912551>
- [12] Gülmez, B. (2026). Artificial intelligence-driven smart grid optimization: A comprehensive review of machine learning, renewable integration, and cybersecurity applications. *Energy Conversion and Management*, 101849. <https://doi.org/10.1016/j.ecmx.2026.101849>
- [13] Haider, Z., Ullah, I., Adel, A., Nasimov, R., & Memon, S. (2025). Artificial Intelligence (AI)-Enabled Unmanned Aerial Vehicle (UAV) Systems for Optimizing User Connectivity in Sixth-Generation (6G) Ubiquitous Networks. *Computers, Materials & Continua*, 86, 1-16. <https://doi.org/10.32604/cmc.2025.071042>
- [14] 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>
- [15] 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>
- [16] 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>
- [17] 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>
- [18] 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>
- [19] Jiao, L., Shao, Y., Sun, L., Liu, F., Yang, S., Ma, W., Li, L., Liu, X., Hou, B., & Zhang, X. (2024). Advanced deep learning models for 6G: Overview, opportunities, and challenges. *IEEE Access*, 12, 133245-133314. <https://doi.org/10.1109/ACCESS.2024.3418900>
- [20] 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>
- [21] 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>
- [22] 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>
- [23] 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>
- [24] 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>
- [25] Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., Niyato, D., Dobre, O., & Poor, H. V. (2021). 6G Internet of Things: A comprehensive survey. *IEEE Internet of Things Journal*, 9(1), 359-383. <https://doi.org/10.1109/JIOT.2021.3103320>
- [26] 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>
- [27] Qian, L. (2022). Joint Multi-domain Resource Allocation and Trajectory Optimization in UAV-assisted M-IoT Networks. *TechRxiv*, 2022(0105). <https://doi.org/10.36227/techrxiv.17704463.v1>
- [28] Richa, Tyagi, A. K., & Nair, M. M. (2026). 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>
- [29] 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>
- [30] 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>
- [31] 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>
- [32] 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>
- [33] 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>
- [34] 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), <https://doi.org/10.1109/ICCES67310.2025.11336962>
- [35] 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>
- [36] 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>
- [37] 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>
- [38] 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>
- [39] 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>