DEEP REINFORCEMENT LEARNING-BASED INTELLIGENT RESOURCE ALLOCATION FOR ENHANCED SPECTRAL EFFICIENCY IN 5G WIRELESS NETWORKS

Miss. DEEPIKA PANDA
Lecturer Stage II
Department of Electronics and Telecommunication Engineering
Govt Polytechnic, Koraput

Mrs. POONAM PANDA
Assistant Professor
Department of Electronics and Communication Engineering
Gandhi Academy of Technology and Engineering (GATE), Berhampur

Miss. KSHIRABDHEE TANAYA DORA
Assistant Professor
Department of Electronics and Communication Engineering
Kalam Institute of Technology, Berhampur

Abstract - The rapid expansion of fifth-generation (5G) wireless communication networks has intensified the demand for intelligent radio resource management strategies capable of improving spectral efficiency, network throughput, and adaptive decision-making under dynamically varying channel conditions. Conventional resource allocation approaches often exhibit limitations in exploiting instantaneous channel variations and optimizing spectrum utilization in heterogeneous multi-user communication environments. To address these challenges, this research proposes a Deep Reinforcement Learning-Based Intelligent Resource Allocation Framework for Enhanced Spectral Efficiency in 5G Wireless Networks, employing the Proximal Policy Optimization (PPO) algorithm for adaptive resource block scheduling. The proposed framework integrates stochastic wireless channel modelling, interference-aware network observations, and a multi-objective reward mechanism incorporating spectral efficiency, user fairness, and interference considerations within a unified reinforcement learning environment. The principal novelty of this study lies in the integration of PPO-driven sequential resource block allocation with a composite performance-oriented reward formulation and systematic benchmarking against conventional scheduling strategies within a reproducible 5G wireless network simulation framework. The developed model was implemented using Python and evaluated in a multi-user downlink communication environment incorporating Rayleigh fading, distance-dependent path loss, and variable channel conditions. Its performance was compared with Round Robin, Random Allocation, and Maximum Signal-to-Interference-plus-Noise Ratio (Max-SINR) scheduling algorithms. The simulation results demonstrated that the proposed PPO-based framework achieved approximately 7.8 bits/s/Hz spectral efficiency and 16.9 Mbps aggregate network throughput, corresponding to spectral efficiency improvements of 98.51%, 87.67%, and 1.96% over Round Robin, Random Allocation, and Max-SINR scheduling, respectively. Furthermore, the training reward convergence characteristics indicated progressive improvement in the learning agent's resource allocation decisions. However, the resulting Jain's fairness index of approximately 0.20 revealed a significant trade-off between aggregate spectral utilization and equitable throughput distribution. The findings demonstrate the feasibility of PPO-based adaptive scheduling for improving spectrum utilization while emphasizing the importance of fairness-aware optimization and scalability in future developments. The proposed framework establishes a computational foundation for investigating intelligent radio resource management in next-generation wireless communication systems.

Index Terms: Fifth-Generation Wireless Networks (5G), Deep Reinforcement Learning (DRL), Proximal Policy Optimization (PPO), Intelligent Resource Allocation, Spectral Efficiency Optimization, Radio Resource Management, Network Throughput, Wireless Channel Modelling, Interference Management, User Fairness, Adaptive Scheduling.

I. INTRODUCTION

The rapid evolution of wireless communication technologies has fundamentally transformed the global telecommunications landscape, creating an unprecedented demand for high-speed data transmission, reliable connectivity, ultra-low latency, and efficient utilization of limited radio frequency resources. The emergence of fifth-generation (5G) wireless communication networks represents a significant technological advancement designed to address the increasingly complex communication requirements of modern digital ecosystems. Unlike previous generations of cellular networks, 5G technology supports a diverse range of applications, including enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC). These capabilities have accelerated the development of intelligent transportation systems, industrial automation, smart cities, autonomous vehicles, remote healthcare, and large-scale Internet of Things (IoT) infrastructures. However, the simultaneous accommodation of heterogeneous communication services, increasing numbers of connected devices, and continuously varying traffic demands introduces substantial challenges in wireless resource management. Consequently, the development of intelligent and adaptive resource allocation mechanisms has become essential for ensuring the efficient

and sustainable operation of next-generation wireless communication networks.

Spectral efficiency is widely recognized as one of the most critical performance indicators in contemporary wireless communication systems because it determines the amount of information that can be transmitted reliably within a given bandwidth. The increasing scarcity of available radio frequency spectrum, combined with the exponential growth of mobile data traffic, has intensified the necessity for communication strategies capable of maximizing spectral utilization without compromising transmission reliability or service quality. In 5G wireless networks, spectral efficiency is influenced by several interdependent factors, including channel quality, transmission power, interference levels, user distribution, bandwidth allocation, modulation techniques, and network topology. The deployment of advanced communication technologies such as massive multiple-input multiple-output (MIMO) as shown in Figure 1, orthogonal frequency-division multiplexing (OFDM) as shown in Figure 2, millimetre-wave communication, beamforming, and dense small-cell architectures has considerably improved the potential for efficient spectrum utilization. Nevertheless, the practical benefits of these technologies depend heavily on the effectiveness of the underlying radio resource management mechanisms. Inefficient allocation of available spectrum and transmission resources can result in excessive inter-user interference, reduced network throughput, uneven service distribution, and underutilization of communication infrastructure.



Figure 1: General architecture of a 5G wireless communication network illustrating the integration of base stations, heterogeneous user equipment, and intelligent radio resource management.

Conventional resource allocation techniques in cellular communication networks primarily rely on predetermined scheduling policies, mathematical optimization frameworks, heuristic algorithms, and rule-based decision-making strategies. Approaches such as proportional fair scheduling, round-robin scheduling, maximum channel quality indicator scheduling, and conventional power allocation methods have been extensively employed to distribute available communication resources among multiple users. Although these approaches provide practical solutions under specific operating conditions, their effectiveness may deteriorate when applied to highly dynamic and heterogeneous 5G network environments. Traditional optimization techniques often require accurate mathematical representations of the network, reliable channel state information, and repeated computational procedures to determine suitable resource allocation decisions. Furthermore, several conventional algorithms exhibit limited adaptability to unpredictable traffic fluctuations, rapidly changing wireless channel conditions, and complex interference patterns. The increasing dimensionality of resource allocation problems, particularly in dense multi-user networks, further complicates the application of conventional optimization approaches. These limitations highlight the necessity for intelligent computational frameworks capable of learning adaptive allocation policies directly from interactions with the wireless environment.Artificial intelligence (AI) and machine learning (ML) have emerged as promising technologies for addressing the computational and operational challenges associated with modern wireless communication networks as shown in Figure 3.

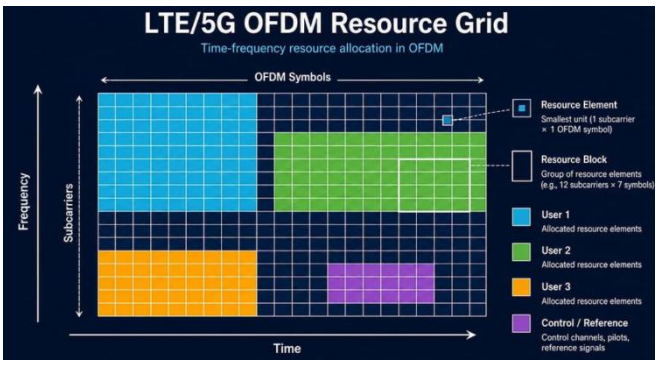


Figure 2: Illustration of radio resource allocation and spectrum utilization in a multi-user 5G wireless communication system.

Their ability to identify complex relationships within multidimensional datasets, predict network behaviour, and support data-driven decision-making has created new opportunities for intelligent radio resource management. Supervised learning algorithms have been investigated for channel estimation, traffic prediction, interference classification, and communication quality assessment, while unsupervised learning approaches have been applied to user clustering, network organization, and resource utilization analysis. However, many conventional machine learning techniques depend on the availability of representative training datasets and may require retraining when network

conditions change substantially. In addition, supervised learning frameworks are not inherently designed to optimize sequential decisions in environments where present allocation choices influence future network performance. These characteristics motivate the investigation of reinforcement learning methodologies that can continuously improve resource allocation decisions through environmental interactions and feedback-driven optimization.

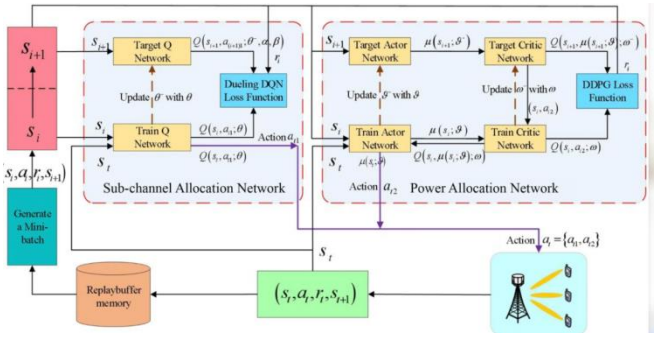


Figure 3: Conceptual framework of deep reinforcement learning-based intelligent resource allocation in 5G wireless networks, illustrating the interaction between the learning agent and the wireless communication environment.

Reinforcement learning (RL) provides a mathematical framework in which an intelligent decision-making agent interacts with a dynamic environment and learns an appropriate control policy by maximizing cumulative rewards. Unlike conventional supervised learning, reinforcement learning does not require predefined optimal decisions for every possible network condition. Instead, the learning agent explores alternative actions, observes their consequences, and gradually identifies strategies that improve long-term system performance. In wireless communication applications, the network environment can be represented through parameters such as channel gain, signal-to-interference-plus-noise ratio (SINR), user traffic demand, interference conditions, and available communication resources. The learning agent can select resource allocation actions involving transmission power, scheduling priorities, bandwidth distribution, or physical resource block assignments. A suitably designed reward mechanism can encourage improvements in spectral efficiency while accounting for fairness, power limitations, and minimum quality-of-service requirements. Nevertheless, conventional reinforcement learning algorithms may encounter significant scalability challenges when the number of network states and possible allocation decisions becomes large, thereby limiting their applicability to realistic multi-user communication scenarios.

Deep reinforcement learning (DRL) addresses several limitations of conventional reinforcement learning by integrating deep neural networks with sequential decision-making algorithms. Through nonlinear function approximation, DRL enables an intelligent agent to learn effective resource allocation policies from high-dimensional network observations without explicitly maintaining a tabular representation of every possible state-action combination. Advanced DRL methodologies, including Deep Q-Networks (DQN), Double Deep Q-Networks (DDQN), Deep Deterministic Policy Gradient (DDPG), and Proximal Policy Optimization (PPO), provide different mechanisms for learning adaptive decision policies under discrete, continuous, and complex action spaces. These algorithms offer considerable potential for wireless resource management because they can accommodate dynamic channel variations, changing user requirements, and nonlinear relationships between resource allocation decisions and communication performance. In particular, DRL-based allocation strategies can be formulated to balance the competing objectives of maximizing network spectral efficiency, maintaining acceptable service quality, minimizing interference, and ensuring equitable access to communication resources.

Despite the growing research interest in DRL-enabled wireless communication systems, the development of reliable and computationally efficient resource allocation mechanisms remains an important research challenge. Existing learning-based approaches may experience slow training convergence, unstable policy updates, excessive exploration costs, and sensitivity to variations in network operating conditions. Furthermore, resource allocation strategies that emphasize aggregate throughput or spectral efficiency without explicitly considering user fairness may preferentially allocate communication resources to users experiencing favourable channel conditions. Such behaviour can reduce the service quality of cell-edge users and create substantial disparities in network performance. Similarly, optimization frameworks that do not adequately incorporate interference constraints and transmission power limitations may produce allocation decisions that are unsuitable for practical wireless systems. The simultaneous consideration of spectral efficiency, interference management, resource utilization, and user fairness therefore represents a significant multi-objective optimization problem. Addressing these challenges requires the development of an adaptive learning framework that can efficiently evaluate changing network conditions and determine suitable resource allocation decisions while respecting operational constraints.

The integration of DRL into 5G radio resource management also introduces important considerations concerning computational complexity, training efficiency, and deployment feasibility. Wireless channel conditions can fluctuate over relatively short timescales because of mobility, multipath propagation, shadowing, and variations in the surrounding electromagnetic environment. Consequently, an intelligent allocation framework must be capable of responding to changing network observations without introducing excessive decision-making delays. Although neural network inference can be computationally efficient after training, the development of an effective policy may require extensive interactions with a simulated network environment. In addition, differences between simulated propagation conditions and actual cellular network behaviour can influence the generalization capability of learned allocation policies. These considerations emphasize the importance of systematic simulation, appropriate channel modelling, controlled training procedures, and comparative performance evaluation. A well-designed DRL framework should demonstrate not only improvements in spectral efficiency but also consistent performance across different user densities, signal quality conditions, and network traffic scenarios.

In this context, Python-based computational modelling provides a flexible and reproducible environment for investigating intelligent resource allocation strategies in 5G wireless networks. Python offers a comprehensive collection of scientific computing, optimization, and machine learning libraries that facilitate wireless channel simulation, neural network development, reinforcement learning implementation, and performance analysis. Libraries such as NumPy, SciPy, PyTorch, TensorFlow, and Matplotlib can be employed to construct a multi-user wireless communication environment, represent channel-dependent resource allocation decisions, train intelligent learning agents, and visualize the resulting network performance. The availability of cloud-based execution platforms such as Google Colab further enables computational experimentation without requiring specialized communication hardware. Through carefully designed simulation scenarios, the effectiveness of DRL-based resource allocation can be evaluated under varying numbers of connected users, available bandwidth, transmission power constraints, interference levels, and channel conditions. Such an approach provides a systematic foundation for examining the relationship between intelligent resource management and achievable spectral efficiency.

The present research investigates a Deep Reinforcement Learning-Based Intelligent Resource Allocation framework for enhancing spectral efficiency in 5G wireless networks. The proposed research direction focuses on formulating the radio resource allocation problem as a sequential decision-making process in which a learning agent observes network conditions and determines appropriate allocation actions to improve communication performance. The framework considers the dynamic relationship among wireless channel quality, user resource requirements, interference, and available communication capacity. By incorporating spectral efficiency into the learning objective and accounting for relevant operational constraints, the approach aims to identify adaptive resource allocation policies that outperform selected conventional scheduling and allocation strategies under the evaluated simulation conditions. The investigation further considers the importance of stable learning behaviour, efficient resource utilization, and equitable service distribution among connected users. Rather than relying exclusively on fixed allocation rules, the intelligent agent is intended to learn allocation strategies that respond to variations in network conditions and evolving communication demands.

The principal objective of this study is to establish a computational framework for investigating how deep reinforcement learning can improve the allocation of limited radio resources in multi-user 5G communication environments. The research seeks to develop an appropriate representation of the wireless network state, define feasible resource allocation actions, and construct a reward function that reflects spectral efficiency while incorporating selected network performance constraints. The learning-based framework will be evaluated against conventional allocation mechanisms through quantitative performance indicators such as aggregate throughput, spectral efficiency, resource utilization, user fairness, and learning convergence characteristics. Particular attention will be given to the behaviour of the intelligent allocation policy under different channel conditions and network loading scenarios. These evaluations are intended to determine the circumstances under which adaptive DRL-based decision-making provides measurable advantages over traditional resource management techniques.

The potential significance of this research extends beyond improvements in a single wireless communication performance metric. An intelligent resource allocation framework capable of adapting to heterogeneous traffic requirements and fluctuating channel conditions could contribute to the broader development of autonomous and software-defined radio access networks. The principles

investigated in this study are also relevant to emerging AI-native network management architectures and future sixth-generation (6G) communication research, where greater network intelligence, distributed decision-making, and efficient spectrum utilization are expected to become increasingly important. By establishing a systematic relationship between deep reinforcement learning and spectral efficiency optimization, the proposed investigation aims to contribute to the understanding of learning-assisted radio resource management in contemporary wireless communication systems. Ultimately, the study seeks to provide a reproducible simulation-based foundation for developing more adaptive, resource-efficient, and performance-aware wireless networks capable of supporting the evolving connectivity requirements of modern digital infrastructure.

II. LITERATURE REVIEW

2.1 Evolution of Resource Allocation Techniques in 5G Wireless Communication Networks

The evolution of fifth-generation (5G) wireless communication networks has introduced substantial advancements in data transmission capabilities, network connectivity, and communication reliability. These improvements have been accompanied by increasing challenges in managing limited radio frequency resources across heterogeneous wireless environments. The efficient allocation of spectrum, bandwidth, transmission power, and scheduling resources has consequently emerged as a fundamental research area in modern cellular communication systems. Earlier investigations into wireless resource management primarily concentrated on conventional optimization methods, including proportional fair scheduling, round-robin scheduling, water-filling power allocation, and channel-aware resource distribution. Such approaches established important foundations for balancing communication throughput, user fairness, and system utilization. However, their performance is frequently influenced by assumptions concerning channel availability, network traffic characteristics, and computational tractability.

The introduction of orthogonal frequency-division multiple access (OFDMA), massive multiple-input multiple-output (MIMO), and dense cellular architectures has increased the complexity of resource allocation problems. In multi-user 5G networks, communication resources must be distributed dynamically according to changing channel conditions, user mobility, interference levels, and service requirements. Traditional allocation mechanisms may struggle to maintain optimal performance when the number of users and possible allocation combinations increases substantially. Consequently, research has progressively shifted toward adaptive optimization frameworks capable of responding to network dynamics and improving spectral utilization through intelligent decision-making.

2.2 Conventional Optimization and Scheduling Approaches

Conventional wireless resource allocation strategies have been extensively investigated to maximize throughput and improve fairness among connected users. Proportional fair scheduling represents a widely adopted approach that attempts to balance instantaneous channel quality against historical user throughput. Similarly, maximum-throughput scheduling prioritizes users experiencing favourable channel conditions, thereby improving aggregate network capacity under certain operating scenarios. Round-robin scheduling distributes communication opportunities among users without directly considering channel variations, offering implementation simplicity but potentially reducing spectral efficiency. These methods demonstrate different trade-offs between throughput maximization, scheduling fairness, and computational complexity.

Mathematical optimization techniques, including convex optimization, Lagrangian relaxation, fractional programming, and successive convex approximation, have also been employed to address resource management problems involving power allocation, bandwidth distribution, and interference constraints. Such approaches can provide mathematically grounded solutions when the optimization structure and required network information are sufficiently well defined. Nevertheless, the presence of discrete allocation decisions, nonlinear interference relationships, and multiple competing objectives can lead to computationally demanding optimization problems. The limitations of conventional techniques become particularly relevant in dense 5G deployments, where rapid changes in wireless propagation conditions and user traffic requirements necessitate frequent allocation decisions.

2.3 Machine Learning-Based Intelligent Resource Management

Machine learning has emerged as a promising alternative for improving wireless network adaptability and reducing dependence on rigid allocation rules. Supervised learning techniques, including artificial neural networks, support vector machines, decision trees, and ensemble learning algorithms, have been investigated for predicting network traffic, estimating channel conditions, and supporting

communication resource management. These approaches can extract nonlinear relationships from historical network observations and provide predictions that assist subsequent allocation decisions. In particular, deep neural networks offer considerable representational capability for modelling complex relationships among channel quality, user demand, interference, and achievable transmission rates.

Despite these advantages, supervised learning-based resource allocation frameworks commonly depend on labelled training examples or reference decisions obtained from existing optimization algorithms. Their effectiveness may deteriorate when network conditions differ significantly from the training distribution. Furthermore, prediction accuracy does not necessarily translate into improved long-term network performance, particularly when resource allocation decisions influence subsequent interference and traffic conditions. These limitations have encouraged the investigation of reinforcement learning, in which resource management policies can be developed through sequential interactions with the wireless environment rather than relying exclusively on predefined optimal allocation labels.

2.4 Deep Reinforcement Learning for Wireless Resource Allocation

Deep reinforcement learning has attracted considerable attention as an intelligent optimization framework capable of addressing sequential resource management problems in complex wireless environments. Luong et al. (2019) presented a comprehensive survey of DRL applications in communications and networking, identifying dynamic spectrum access, transmission rate control, resource sharing, and network optimization as important areas for learning-based decision-making. Their investigation established that deep neural networks can improve the applicability of reinforcement learning to environments characterized by large state and action spaces. The study also emphasized challenges associated with computational requirements, learning stability, and the practical implementation of DRL algorithms in communication networks.

Alwarafy et al. (2021) systematically examined DRL-based radio resource allocation and management in heterogeneous wireless networks. Their research classified existing approaches according to the resources being managed, network architectures, and learning algorithms. The investigation highlighted the potential of value-based and policy-based DRL methods for addressing resource management challenges involving network heterogeneity, complex operating conditions, and dynamic communication demands. The authors also identified unresolved issues concerning scalability, training complexity, and generalization across different wireless environments. These findings provide an important foundation for investigating adaptive resource allocation strategies in multi-user 5G networks as shown in Figure 4.

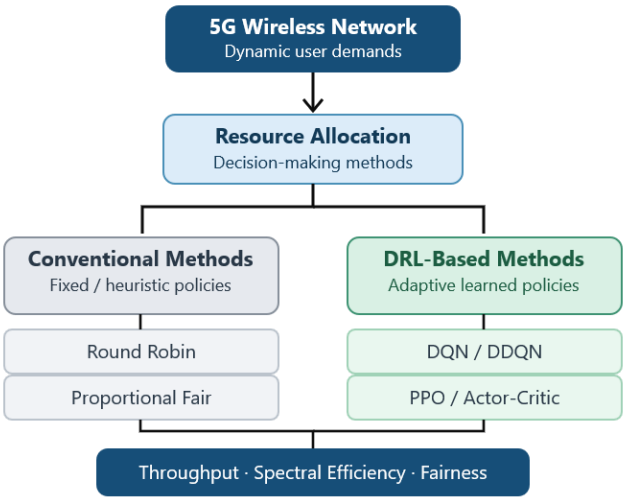


Figure 4: Classification of conventional and deep reinforcement learning-based radio resource allocation techniques for 5G wireless communication networks. Original conceptual illustration based on the reviewed approaches.

2.5 Deep Q-Networks and Value-Based Resource Allocation

Deep Q-Networks represent an important category of value-based reinforcement learning algorithms in which neural networks approximate the expected cumulative reward associated with available decisions. Their application to wireless resource allocation enables the learning agent to evaluate alternative actions based on observed channel conditions, user requirements, and network performance indicators. In a discrete allocation environment, a DQN agent can learn to select suitable resource blocks, scheduling decisions, or predefined transmission power levels. Experience replay and target network mechanisms are commonly employed to reduce correlations between training samples and improve learning stability.

However, conventional DQN approaches encounter practical limitations when resource allocation involves large combinations of users, channels, and power levels. The number of possible allocation actions can increase rapidly as network dimensions expand, making direct action-value evaluation computationally expensive. Double DQN and related extensions can reduce certain learning biases, but they do not automatically resolve the underlying action-space scalability problem. Consequently, value-based DRL

methods are particularly attractive for discretized allocation scenarios, while more complex resource management tasks may require structured action representations or alternative policy optimization techniques.

2.6 Policy-Based Deep Reinforcement Learning and Adaptive Optimization

Policy-based reinforcement learning algorithms have received increasing attention because of their ability to optimize parameterized decision policies directly. Actor-critic methods combine a policy network responsible for selecting actions with a value-estimation mechanism that evaluates the expected quality of decisions. Algorithms such as Deep Deterministic Policy Gradient and Proximal Policy Optimization provide useful alternatives to conventional value-based approaches, particularly when resource allocation includes continuous control variables or complex decision policies. PPO employs a constrained policy-update objective designed to limit excessively large changes between successive policy iterations, which can contribute to more stable training under suitable conditions.

In the context of wireless communication, policy-based learning can support adaptive transmission power control, bandwidth distribution, and resource scheduling. Nevertheless, the effectiveness of these algorithms depends on appropriate state representations, reward formulations, neural network architectures, and training configurations. Policy-gradient methods can also exhibit substantial sample requirements, while continuous action outputs may need additional processing to satisfy discrete resource block assignments and practical transmission constraints. Therefore, the selection of a suitable DRL algorithm should be guided by the mathematical structure of the resource allocation problem rather than by algorithm popularity alone.

2.7 Spectral Efficiency and Interference-Aware Resource Optimization

Spectral efficiency optimization represents a central objective of wireless resource management because it directly influences the amount of information transmitted per unit bandwidth. In multi-user 5G systems, achievable spectral efficiency depends on the interaction among channel gain, transmission power, interference, noise, and allocation decisions. Increasing transmission power may improve the received signal quality of an individual user while simultaneously increasing interference experienced by other users when resources are reused. Similarly, allocating resources preferentially to users with favourable propagation conditions may increase aggregate spectral efficiency but reduce fairness and service availability for users experiencing weaker channels.

Research on intelligent radio resource management has therefore emphasized the importance of considering multiple performance objectives rather than maximizing throughput in isolation. Seguin et al. (2023), in their review of DRL-based downlink scheduling for 5G and beyond networks, examined fine-grained approaches involving the allocation of resource blocks to user equipment. Their review highlights the significance of heterogeneous quality-of-service requirements and the growing interest in learning-based scheduling strategies.

For spectral-efficiency-oriented resource allocation, the reward function can be formulated to incorporate aggregate achievable data rate, bandwidth utilization, fairness, and penalties for interference or constraint violations. Such formulations allow the learning process to reflect competing communication objectives. However, the effectiveness of a multi-objective reward depends strongly on the relative weighting of its components and the physical assumptions underlying the simulated wireless environment.

2.8 Challenges in DRL-Based 5G Resource Management

Although deep reinforcement learning provides a promising foundation for intelligent wireless resource management, several technical challenges remain unresolved. Khani et al. (2024) reviewed DRL-based resource allocation in 5G cloud radio access networks and identified scalability, convergence, and fairness as important considerations in the development of practical learning-based allocation mechanisms. Their work demonstrates the continuing research interest in autonomous network management while emphasizing the difficulties associated with deploying DRL in complex cellular environments.

One major challenge concerns the relationship between network complexity and learning efficiency. As the number of connected users, available resource blocks, and transmission configurations increases, the state-action space becomes substantially larger. This expansion can increase training requirements and complicate policy convergence. Another important issue is the nonstationary nature of wireless environments, where user mobility, fluctuating channel quality, and variable traffic demand may alter the relationship between allocation decisions and achieved rewards. Furthermore, the use of approximate neural network policies introduces concerns regarding reproducibility, generalization, constraint satisfaction, and computational overhead. These challenges indicate that improvements in spectral efficiency must be evaluated

alongside algorithmic stability, user fairness, and operational feasibility.

2.9 Research Gaps Identified from Existing Literature

A critical examination of the available literature indicates that substantial progress has been achieved in applying deep reinforcement learning to wireless resource management. Nevertheless, the simultaneous optimization of spectral efficiency, interference management, user fairness, and computational efficiency remains an important research direction. Existing investigations have explored individual allocation objectives and various learning algorithms; however, the performance of such approaches can vary considerably with network topology, channel conditions, traffic characteristics, and reward-function design. The absence of a universally effective allocation policy creates opportunities for developing application-specific learning frameworks capable of adapting to dynamic multi-user environments.

Another important research gap concerns the trade-off between spectral efficiency maximization and equitable resource distribution. Policies designed primarily to maximize aggregate transmission rates may disproportionately favour users experiencing strong wireless channels, while fairness-oriented strategies can reduce overall network throughput. In addition, complex resource allocation environments may require extensive DRL training and introduce difficulties in maintaining stable policy performance under changing operating conditions. These observations motivate the investigation of reward functions that explicitly balance spectral efficiency with fairness and interference constraints, together with systematic comparisons against conventional resource allocation strategies.

2.10 Motivation for the Proposed Research

The preceding literature establishes deep reinforcement learning as a promising computational approach for adaptive radio resource management while revealing persistent challenges associated with multi-objective optimization, training stability, and network scalability. Motivated by these observations, the present study proposes to investigate an intelligent DRL-based resource allocation framework for enhancing spectral efficiency in 5G wireless communication networks. The proposed approach will model the interaction between network observations, allocation decisions, and resulting communication performance through a reinforcement learning environment. The principal emphasis will be on learning resource allocation policies that respond dynamically to changing channel conditions and user requirements while accounting for interference, available bandwidth, and fairness considerations.

The study will employ Python-based simulation to examine the effectiveness of the proposed learning framework under controlled network conditions. Performance will be evaluated using spectral efficiency, network throughput, fairness, resource utilization, and learning convergence characteristics. Comparative analysis against conventional scheduling and allocation techniques will be used to determine whether the DRL framework provides measurable improvements. The intended research contribution is therefore the development and evaluation of a reproducible, constraint-aware, and performance-oriented intelligent resource allocation methodology for multi-user 5G wireless networks.

III. PROBLEM STATEMENT

The rapid expansion of fifth-generation (5G) wireless communication networks has introduced significant challenges in achieving efficient radio resource utilization while accommodating the increasing demands of heterogeneous communication services. The widespread deployment of high-density cellular infrastructures, massive multiple-input multiple-output (MIMO) systems, Internet of Things (IoT) devices, and bandwidth-intensive applications has substantially increased the complexity of wireless resource management. Although 5G technology incorporates advanced communication mechanisms to improve network capacity, transmission reliability, and connectivity, the available radio frequency spectrum remains inherently limited. Consequently, the efficient allocation of bandwidth, transmission power, and physical resource blocks among multiple users has become a critical requirement for maximizing spectral efficiency and maintaining acceptable communication performance. The principal challenge lies in determining appropriate resource allocation decisions under dynamically varying wireless channel conditions, interference levels, traffic demands, and user distribution patterns without introducing excessive computational complexity or compromising network fairness.

In practical 5G communication environments, spectral efficiency is strongly influenced by the interaction among channel state information, signal-to-interference-plus-noise ratio (SINR), resource block availability, transmission power, and inter-user interference. These parameters exhibit nonlinear and time-varying characteristics, making the resource allocation problem increasingly difficult to solve using static scheduling policies and conventional mathematical optimization techniques. Traditional

approaches, including round-robin scheduling, proportional fair scheduling, and channel-quality-based allocation, generally operate according to predefined decision rules that may not adequately adapt to rapidly changing network conditions. Although mathematical optimization algorithms can provide effective allocation solutions under specific assumptions, their computational requirements may increase considerably in large-scale multi-user systems involving discrete resource assignments and nonlinear interference constraints. Furthermore, conventional allocation mechanisms may prioritize either aggregate throughput or user fairness, making it difficult to achieve a satisfactory balance between spectral efficiency, interference mitigation, and equitable resource distribution.

The complexity of the resource allocation problem becomes particularly significant when multiple users compete for limited communication resources while experiencing different propagation conditions and quality-of-service requirements. Users located near a base station may benefit from stronger received signal power and higher achievable transmission rates, whereas cell-edge users frequently experience weaker channel conditions and increased interference. Resource allocation policies designed exclusively to maximize instantaneous throughput may repeatedly Favor users with favourable channel characteristics, leading to inefficient service distribution and reduced fairness. Conversely, allocation mechanisms that distribute resources uniformly without considering channel quality may fail to exploit available transmission opportunities, resulting in lower aggregate spectral efficiency. Therefore, an effective resource management strategy must consider the competing objectives of maximizing spectral utilization, maintaining communication reliability, controlling interference, and ensuring appropriate access to available network resources.

Recent developments in artificial intelligence and machine learning have created opportunities for addressing complex wireless optimization problems through data-driven decision-making. However, conventional supervised learning approaches often depend on representative training datasets and predefined target decisions, which may limit their adaptability to previously unseen communication scenarios. Reinforcement learning provides an alternative by allowing an intelligent agent to improve allocation decisions through repeated interactions with the network environment. Nevertheless, conventional reinforcement learning methods encounter scalability limitations when applied to high-dimensional wireless systems containing numerous users, resource blocks, and possible transmission configurations. The exponential growth of feasible allocation combinations can make explicit state-action representations computationally impractical. Deep reinforcement learning addresses this limitation by employing neural networks to approximate decision policies or action-value functions, thereby enabling adaptive resource management in complex environments. Despite these advantages, the effectiveness of DRL-based resource allocation depends on appropriate state representation, action-space design, reward formulation, training stability, and the ability to satisfy communication constraints.

A major unresolved challenge is the development of an intelligent resource allocation framework that can improve spectral efficiency while maintaining fairness and adapting to dynamic network conditions. DRL agents trained primarily to maximize aggregate throughput may learn policies that favour users with stronger channels, whereas reward functions containing several competing performance indicators may produce unstable or inefficient learning behaviour when their relative contributions are not properly balanced. Furthermore, wireless resource allocation decisions must satisfy practical constraints associated with finite bandwidth, available physical resource blocks, transmission power budgets, and interference limitations. The simultaneous consideration of these requirements introduces a constrained, nonlinear, and potentially combinatorial optimization problem. Consequently, the development of a computationally manageable learning framework capable of identifying effective allocation policies across different network conditions remains an important research objective.

The present research addresses this problem by investigating a Deep Reinforcement Learning-Based Intelligent Resource Allocation Framework for Enhanced Spectral Efficiency in 5G Wireless Networks. The proposed framework is intended to model a multi-user 5G communication environment in which an intelligent learning agent observes relevant network parameters, including channel quality, interference conditions, user demands, and available communication resources. Based on these observations, the agent determines suitable resource allocation actions with the objective of improving spectral efficiency while considering network fairness and operational constraints. A reward-driven learning mechanism will be formulated to guide the optimization process, enabling the agent to identify allocation strategies that respond to variations in the wireless environment. The framework will be developed and evaluated using Python-based computational simulation, providing a reproducible platform for examining intelligent resource management under different communication scenarios.

The central research problem can therefore be expressed as the determination of an adaptive resource allocation policy that maximizes the aggregate spectral efficiency of a multi-user 5G wireless communication system while satisfying available bandwidth, transmission power, interference, and user service constraints. The investigation will examine whether a DRL-based decision-making strategy can achieve improved spectral utilization compared with conventional allocation mechanisms under equivalent network conditions. Performance will be assessed using quantitative indicators such as aggregate spectral efficiency, network throughput, user fairness, resource utilization, and training convergence. Particular emphasis will be placed on evaluating the effectiveness of the learned allocation policy under varying user densities, channel quality conditions, and network loading scenarios.

By addressing these challenges, the proposed research seeks to establish a systematic relationship between intelligent sequential decision-making and efficient radio resource utilization in 5G wireless communication networks. The anticipated contribution lies in the formulation, implementation, and comparative evaluation of an adaptive resource allocation methodology that integrates deep reinforcement learning with spectral-efficiency-oriented network optimization. The findings are expected to provide insights into the feasibility and limitations of learning-assisted resource management and support further investigations into autonomous wireless network optimization, intelligent radio access networks, and future AI-enabled communication systems.

IV. MATHEMATICAL MODELLING

The proposed Deep Reinforcement Learning (DRL)-based intelligent resource allocation framework considers a multi-user 5G wireless communication network consisting of a base station serving multiple user equipment (UE) devices through a finite number of orthogonal physical resource blocks (RBs). The mathematical model aims to maximize spectral efficiency by dynamically optimizing resource block assignments and transmission power while accounting for channel conditions, interference, and user fairness. The resource allocation problem is formulated as a Markov Decision Process (MDP), enabling a DRL agent to learn an adaptive allocation policy through continuous interaction with the simulated wireless environment.

4.1 Wireless Channel Model

The wireless communication channel is modelled using distance-dependent path loss and Rayleigh fading to represent large-scale attenuation and small-scale multipath propagation. The effective channel power gain between the base station and user \(k\) on resource block \(n\) at time \(t\) is expressed as

$$g_{k,n}(t) = |h_{k,n}(t)|^2 d_k^{-\alpha} \quad (1)$$

where $h_{k,n}(t)$ represents the complex Rayleigh fading coefficient, d_k denotes the distance between the base station and user k, and α represents the path-loss exponent.

4.2 Signal-to-Interference-plus-Noise Ratio

The signal-to-interference-plus-noise ratio (SINR) determines the quality of the received wireless signal and directly influences the achievable data transmission rate. For user k operating on resource block \(n\), the SINR is defined as

$$\gamma_{k,n}(t) = \frac{p_{k,n}(t) g_{k,n}(t)}{I_{k,n}(t) + \sigma_n^2} \quad (2)$$

where $p_{k,n}(t)$ is the allocated transmission power, $I_{k,n}(t)$ represents interference from other transmissions, and σ_n^2 denotes the noise power associated with the resource block.

For the considered single-cell downlink OFDMA system, each resource block is assigned to at most one user. Consequently, intracell interference on an assigned resource block is avoided under ideal orthogonal scheduling, while $I_{k,n}(t)$ can represent interference from neighbouring cells.

4.3 Achievable Data Rate

The achievable data rate for user k is estimated using the Shannon capacity expression:

$$R_k(t) = \sum_{n=1}^{N} x_{k,n}(t) B_n \log_2 (1 + \gamma_{k,n}(t)) \quad (3)$$

where B_n represents the bandwidth of resource block n, N is the total number of available resource blocks, and $x_t \; in \{0,1\}$ indicates whether resource block n is assigned to user k. This expression represents an idealized achievable-rate estimate rather than the exact throughput of a practical 5G modulation and coding scheme.

4.4 Spectral Efficiency

Spectral efficiency is defined as the total achievable network data rate per unit of allocated system bandwidth:

$$\eta_{SE}(t) = \frac{\sum_{k=1}^{K} R_k(t)}{\sum_{n=1}^{N} B_n} \quad (4)$$

where K denotes the total number of users. Spectral efficiency is measured in bits/s/Hz and serves as the principal optimization objective of the proposed framework.

4.5 Resource Allocation Optimization

The resource allocation problem is formulated to maximize the expected aggregate spectral efficiency subject to available communication resources and transmission power limitations.

$$\begin{aligned} \max_{\mathbf{x},\mathbf{p}} \quad & \mathbb{E}\left[\eta_{\mathrm{SE}}(t)\right] \\ \text{subject to} \quad & \sum_{k=1}^{K} x_{k,n}(t) \le 1, \quad \forall n, \\ & \sum_{k=1}^{K}\sum_{n=1}^{N} x_{k,n}(t) p_{k,n}(t) \le P_{\max}, \\ & x_{k,n}(t) \in \{0,1\}, \quad 0 \le p_{k,n}(t) \le P_{\max}. \end{aligned} \tag{5}$$

Here, P_{max} represents the maximum transmission power budget of the base station. The first constraint prevents multiple users from simultaneously occupying the same resource block within the considered cell, while the second ensures that total transmission power remains within the permissible limit.

4.6 Markov Decision Process Formulation

The intelligent resource allocation problem is represented as an MDP defined by the tupleS,A,P,r, γ, where S represents the state space, A represents the action space, P denotes the state transition probabilities, r is the reward function, and γ is the discount factor.

The network state is represented as

$$s_t = [\mathbf{G}_t, \mathbf{\Gamma}_t, \mathbf{D}_t, \mathbf{Q}_t] \tag{6}$$

where G_t represents channel gains, γ_t denotes SINR observations, $D_t$contains user traffic demands, and Q_t represents the user queue states. The DRL agent selects an action a_t consisting of feasible resource block assignments and transmission power decisions.

4.7 Reward Function

The reward function guides the DRL agent toward resource allocation decisions that improve spectral efficiency while considering fairness and interference.

$$r_t = w_1\eta_{\mathrm{SE}}(t) + w_2 J(t) - w_3 I_{\mathrm{norm}}(t) \tag{7}$$

where w_1,w_2,w_3 are nonnegative weighting coefficients, J(t) denotes Jain's fairness index, and I_{norm}(t) represents normalized interference. The reward components should be scaled appropriately so that their relative contributions are controlled by the selected weights.

Jain's fairness index is calculated as

$$J(t) = \frac{\left(\sum_{k=1}^{K} R_k(t)\right)^2}{K\sum_{k=1}^{K} R_k^2(t)} \tag{8}$$

The fairness index ranges from 1/K to 1 when at least one user has a positive rate, with higher values indicating more equitable throughput distribution. When all user rates are zero, fairness is defined as zero for the simulation.

4.8 Deep Reinforcement Learning Objective

The objective of the DRL agent is to learn an optimal policy π_θ that maximizes the expected cumulative discounted reward:

$$\pi_\theta^* = \arg\max_{\pi_\theta} \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{T-1} \gamma^t r_t\right] \tag{9}$$

where θ represents the trainable neural network parameters, T is the simulation horizon, and $\gamma \in [0,1)$ controls the relative importance of future rewards.

For a Proximal Policy Optimization (PPO)-based implementation, the clipped surrogate objective is expressed as

$$L^{\mathrm{CLIP}}(\theta) = \mathbb{E}_t\left[\min\left(\rho_t(\theta)\hat{A}_t,\ \mathrm{clip}\left(\rho_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t\right)\right] \tag{10}$$

The proposed mathematical formulation provides the foundation for developing a Python-based DRL resource allocation framework. The wireless channel model, SINR estimation, achievable data rate, and spectral efficiency equations characterize the communication environment, while the MDP formulation, reward function, and PPO objective establish the intelligent decision-making mechanism. Together, these equations enable systematic evaluation of spectral efficiency, throughput, fairness, interference, and learning convergence under different simulated 5G network conditions.

Implementation note: Equations (1)– (10) are suitable for a simplified single-cell 5G downlink OFDMA simulation. The Python implementation should use the same channel, interference, scheduling, and power constraints to ensure consistency between the mathematical model and experimental results.

V. MODEL IMPLEMENTATION

The proposed Deep Reinforcement Learning-Based Intelligent Resource Allocation framework was implemented using Python to investigate the effectiveness of adaptive radio resource management in fifth-generation (5G) wireless communication networks. The implementation was designed to establish a computational environment capable of representing multi-user wireless communication, dynamic channel variations, resource allocation decisions, and the corresponding network performance characteristics. The framework integrates wireless communication modelling with a deep reinforcement learning algorithm to develop an intelligent decision-making mechanism for improving spectral efficiency under changing network conditions. Google Colab was selected as the primary computational platform because of its accessibility, compatibility with modern machine learning libraries, and ability to execute neural network training and numerical simulations within a unified environment. The implementation employed NumPy for numerical operations and wireless channel calculations, Gymnasium for constructing the reinforcement learning environment, Stable-Baselines3 for implementing the Proximal Policy Optimization algorithm, Pandas for organizing experimental results, and Matplotlib for generating graphical representations of network performance.

The initial stage of implementation involved developing a simulated 5G cellular communication environment consisting of a single base station serving multiple wireless users through a limited collection of orthogonal radio resource blocks. The simulated network was configured to represent a downlink communication scenario in which the base station distributes available transmission resources among users experiencing different propagation conditions. Eight wireless users and twelve resource blocks were considered in the primary experimental configuration, with each resource block assigned a bandwidth of 180 kHz. The total transmission power of the base station was maintained at 30 dBm, while the propagation environment incorporated distance-dependent path loss and Rayleigh fading to reproduce essential characteristics of wireless signal attenuation and multipath propagation. User distances were randomly generated within the considered cellular coverage region, allowing the simulation to represent variations in received signal strength across different user locations. The resource block configuration provided a simplified orthogonal frequency-division multiple access environment suitable for examining the influence of intelligent scheduling on spectral utilization. Although the model incorporates important features of 5G radio resource management, it does not attempt to reproduce the complete physical-layer processing or protocol stack specified by the 5G New Radio standard.

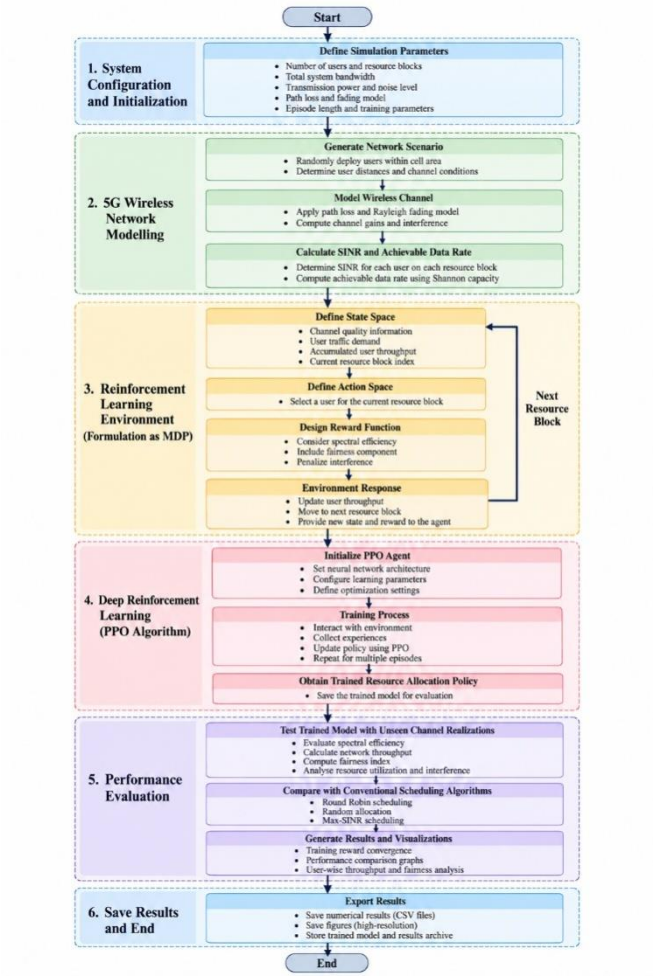


Figure 5: Academic flowchart illustrating the implementation of the proposed deep reinforcement learning-based intelligent resource allocation framework for enhanced spectral efficiency in 5G wireless networks, including network initialization, wireless channel modelling, Markov decision process formulation, PPO training, performance evaluation, and results generation.

The wireless propagation environment was developed to represent the influence of channel fading, propagation distance, background noise, and external interference on communication performance. Rayleigh fading was employed to model small-scale signal fluctuations arising from multipath propagation, while distance-dependent attenuation represented the reduction in received signal strength as the separation between the transmitting base station and receiving users increased. The simulation also incorporated a simplified external interference component to examine the effect of unwanted signals on the quality of wireless reception. Thermal noise was estimated using the configured resource block bandwidth and receiver noise characteristics. These propagation components were

combined to calculate the signal-to-interference-plus-noise ratio associated with individual users and resource blocks. The resulting channel quality information was subsequently used to determine achievable transmission rates and evaluate the effectiveness of alternative resource allocation decisions. By generating different wireless channel realizations across simulation episodes, the implementation provided a controlled environment for investigating adaptive scheduling under varying propagation conditions.

Following the development of the wireless communication environment, a custom reinforcement learning environment was constructed using the Gymnasium framework. This environment established the interaction mechanism between the intelligent resource allocation agent and the simulated 5G network. The environment was designed to provide the learning agent with observations describing the current communication conditions and to evaluate the consequences of its allocation decisions. The observation structure incorporated normalized channel quality information, relative user traffic demands, accumulated user throughput, and the position of the current resource block within the allocation sequence. These observations enabled the learning agent to consider both the immediate quality of available wireless channels and the resource distribution achieved during earlier scheduling decisions. Numerical normalization was applied to maintain a consistent observation range and support stable neural network processing. The resulting environment provided a structured interface through which the reinforcement learning algorithm could interact with the wireless network without requiring physical communication hardware.

The resource allocation process was implemented as a sequential decision-making procedure in which the learning agent selected a user for each available resource block. At the beginning of an allocation episode, the environment generated the user locations, wireless channel conditions, interference characteristics, and relative traffic demands. The learning agent subsequently received an observation describing the current network state and selected the user to whom the next resource block would be assigned. Following each decision, the environment calculated the achievable data rate for the selected user and updated the accumulated throughput distribution. The agent then received a numerical reward reflecting the quality of the allocation decision before proceeding to the next available resource block. This interaction continued until all resource blocks had been assigned, at which point the episode was terminated and the network performance indicators were recorded. A new channel realization was generated for the following episode, allowing the learning process to experience different communication conditions and allocation outcomes.

The intelligent decision-making mechanism was developed using the Proximal Policy Optimization algorithm, a policy-gradient-based deep reinforcement learning technique designed to improve learning stability through controlled policy updates. PPO was selected because of its suitability for sequential decision-making problems involving discrete actions and nonlinear relationships between observations and outcomes. The implementation employed a multilayer perceptron policy architecture provided by the Stable-Baselines3 framework. The neural network received the normalized wireless network observations and generated a probability distribution over the available user selection actions. During training, the agent explored alternative allocation decisions and progressively adjusted its neural network parameters according to the rewards obtained from the simulated environment. The PPO training procedure incorporated advantage estimation, policy clipping, minibatch optimization, and repeated neural network updates to support the development of an adaptive scheduling policy. The implementation used a learning rate of 0.0003, a discount factor of 0.99, a minibatch size of 64, and a policy clipping threshold of 0.2. These settings established an initial training configuration rather than representing universally optimal hyperparameter values.

A reward-based optimization mechanism was incorporated to guide the intelligent agent toward resource allocation decisions that considered spectral efficiency, user fairness, and interference. The reward formulation combined the incremental contribution of each allocation decision to network spectral efficiency with a fairness-related component and an interference penalty. Spectral efficiency was included as the principal performance objective because it reflects the effectiveness with which the available communication bandwidth is utilized. The fairness component encouraged the agent to consider the distribution of achieved transmission rates among connected users rather than concentrating exclusively on users experiencing favourable channel conditions. The interference penalty introduced an additional consideration related to the external interference experienced by the selected user. These reward components were combined using predefined weighting coefficients to establish a multi-objective learning signal. The resulting reward mechanism was intended to encourage the development of allocation policies that balance efficient spectrum utilization with equitable resource distribution and interference-aware decision-making. The effectiveness of this balance was assessed through the subsequent

performance evaluation rather than assumed from the reward formulation alone.

The training stage was conducted through repeated interactions between the PPO agent and the simulated wireless communication environment. A total of 30,000 training time steps was specified for the initial experiment, during which the learning agent repeatedly observed network conditions, selected resource allocation actions, and received performance-based feedback. Each allocation episode consisted of twelve sequential resource block assignment decisions, corresponding to the number of available resource blocks in the simulated network. The training process was configured to update the policy using collected experience and to progressively improve the expected cumulative reward. A dedicated callback mechanism recorded the cumulative reward obtained during completed training episodes, enabling subsequent examination of learning behaviour and convergence characteristics. A fixed random seed was employed to improve the reproducibility of the computational experiment. After training, the learned PPO model was saved for further evaluation, allowing the trained policy to be reused without repeating the entire optimization procedure.

To assess the effectiveness of the proposed intelligent allocation strategy, three conventional scheduling approaches were implemented as comparative benchmarks. The Round Robin algorithm assigned resource blocks to users in a predetermined cyclic sequence, providing a straightforward reference method that distributes scheduling opportunities without directly prioritizing channel quality. The Random Allocation method selected users randomly for resource block assignments, representing a non-optimized scheduling strategy. The Maximum-SINR algorithm assigned each resource block to the user experiencing the highest instantaneous signal-to-interference-plus-noise ratio, thereby providing a channel-aware benchmark oriented toward maximizing instantaneous achievable transmission rates. These comparison methods were implemented within the same wireless communication environment to maintain consistency in the propagation assumptions, resource availability, transmission power, and performance calculations. The inclusion of multiple baselines enabled the investigation of whether the learned PPO policy provided measurable benefits over fixed, random, and channel-aware resource allocation strategies.

The evaluation stage was performed using independently generated wireless channel realizations that were not used during the initial training process. Twenty evaluation episodes were considered for the principal performance comparison, with identical channel realizations supplied to each scheduling algorithm within the corresponding evaluation episode. This approach reduced the influence of differences in randomly generated propagation conditions and supported a more consistent comparison of allocation performance. The trained PPO policy was evaluated using deterministic action selection, while the baseline methods followed their respective scheduling rules. During each evaluation episode, the environment recorded the achieved network throughput, aggregate spectral efficiency, and Jain's fairness index. These performance indicators were selected because they characterize complementary aspects of wireless resource management, including overall transmission capacity, bandwidth utilization, and equitable service distribution. The collected observations were organized into structured datasets for subsequent statistical analysis and graphical presentation.

The performance analysis was designed to examine the behaviour of the proposed framework from several perspectives. Training reward convergence was investigated to assess whether the learning agent exhibited improving and stabilizing reward characteristics during repeated interactions with the wireless environment. Spectral efficiency was evaluated to determine the amount of achievable information transmission per unit bandwidth, while network throughput was examined to quantify the aggregate communication capacity supported by each scheduling method. Jain's fairness index was employed to evaluate the distribution of transmission rates among connected users and identify potential differences between throughput-oriented and fairness-aware allocation strategies. Comparative bar charts were generated to illustrate the performance of PPO, Round Robin, Random Allocation, and Maximum-SINR scheduling under equivalent network conditions. The graphical analysis provided a systematic basis for interpreting the relative advantages and limitations of the learning-based resource allocation approach.

An additional simulation experiment was incorporated to investigate the relationship between the number of connected users and achievable spectral efficiency. The number of users was varied across selected network configurations while the number of available resource blocks and the underlying propagation assumptions were maintained. This experiment was intended to illustrate how increasing competition for limited communication resources influences network performance. However, the implemented PPO neural network was trained using a fixed observation structure corresponding to eight users. Consequently, the trained policy was directly applicable only to the original

eight-user configuration, while the conventional scheduling algorithms could be evaluated across the wider range of user counts. The resulting scalability analysis therefore provided a preliminary examination of conventional scheduling behaviour rather than a complete comparison of PPO scalability. A comprehensive investigation of the learning-based approach under different network sizes would require separately trained PPO models or a neural network architecture capable of accommodating variable numbers of users.

The implementation also incorporated automated data management and visualization procedures to support reproducible computational experimentation. Numerical evaluation results, training reward histories, algorithmic performance summaries, and user-density observations were exported as comma-separated value files. Graphical outputs were generated using Matplotlib and saved at a resolution of 300 dots per inch to support inclusion in technical reports and research manuscripts. The trained PPO model and experimental outputs were stored in a dedicated results directory, and the generated files were packaged into a compressed archive for convenient retrieval from Google Colab. These procedures established a consistent workflow for conducting simulations, preserving experimental observations, and preparing figures for subsequent interpretation. The availability of numerical datasets also enables additional statistical evaluation, sensitivity analysis, and comparison with alternative learning algorithms.

The implemented framework represents a simplified but structured approach to investigating deep reinforcement learning for intelligent resource allocation in 5G wireless communication networks. Its principal computational contribution lies in integrating a stochastic wireless channel environment, sequential resource block scheduling, multi-objective reward evaluation, and PPO-based policy learning within a unified Python simulation platform. The framework enables controlled comparisons between adaptive learning-based scheduling and conventional allocation strategies while providing measurable indicators of spectral efficiency, throughput, fairness, and training behaviour. The current implementation focuses on intelligent resource block assignment with equal transmission power distributed across the available resource blocks; independent power optimization, user mobility, detailed traffic queues, and complete 5G protocol-level modelling are outside the scope of the initial simulation. Nevertheless, the modular design provides a foundation for future extensions involving joint power and bandwidth allocation, more realistic interference environments, multi-cell coordination, and advanced deep reinforcement learning algorithms. Through systematic computational evaluation, the proposed implementation aims to establish the conditions under which learning-assisted radio resource management can contribute to more adaptive and efficient spectrum utilization in 5G wireless communication systems.

VI. RESULTS AND DISCUSSION

The performance of the proposed Deep Reinforcement Learning-Based Intelligent Resource Allocation framework was investigated through a series of Python-based simulation experiments designed to evaluate its effectiveness in improving spectral efficiency and network throughput in fifth-generation (5G) wireless communication networks. The experimental evaluation was conducted using a simulated multi-user downlink communication environment incorporating Rayleigh fading, distance-dependent path loss, variable channel conditions, and external interference. The Proximal Policy Optimization (PPO) algorithm was employed to develop an adaptive resource allocation policy, and its performance was compared with three conventional scheduling strategies, namely Round Robin, Random Allocation, and Maximum Signal-to-Interference-plus-Noise Ratio (Max-SINR) scheduling. The evaluation considered training reward convergence, spectral efficiency, aggregate network throughput, user fairness, and the influence of increasing user density on resource utilization. The obtained results provide insights into the effectiveness of reinforcement learning for adaptive wireless resource allocation while highlighting the trade-offs between maximizing communication capacity and maintaining equitable resource distribution.

6.1 PPO Training Reward Convergence Analysis

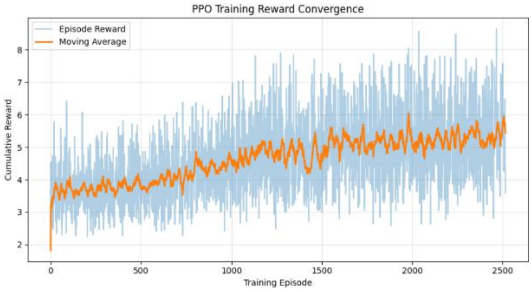


Figure 6: PPO training reward convergence

The training reward convergence characteristics of the proposed PPO-based intelligent resource allocation framework are illustrated in Figure 5. The graph represents the variation in cumulative reward over approximately 2,500 training episodes, together with a moving-average curve that provides a clearer representation of the overall learning

trend. During the initial training phase, the cumulative reward exhibited considerable fluctuations, with the moving average remaining approximately between 3.5 and 4.0. This behaviour indicates that the learning agent was exploring different resource allocation decisions and had not yet developed a consistently effective scheduling policy. The fluctuations observed during this stage are consistent with the stochastic nature of the simulated wireless environment, where changes in channel quality, interference conditions, and user requirements influence the reward associated with individual allocation decisions.

As training progressed, the moving-average reward demonstrated a gradual upward trend, particularly after approximately 600 training episodes. Between 800 and 1,500 episodes, the moving-average reward increased from approximately 4.3 to nearly 5.0, indicating that the agent was learning allocation decisions associated with improved cumulative reward. During the later stages of training, the moving average generally remained above 5.0 and approached approximately 5.5 toward the end of the simulation. Individual episode rewards continued to exhibit substantial variability, with occasional values exceeding 8.0 and other episodes producing considerably lower rewards. These variations reflect the changing propagation conditions and the stochastic nature of the learning environment.

The overall improvement in the moving-average reward suggests that the PPO algorithm successfully learned a policy that achieved higher rewards than those obtained during the early training phase. The results also demonstrate the ability of the neural network-based agent to adjust its scheduling behaviour through repeated interactions with the wireless environment. However, the continued fluctuations in the reward curve indicate that complete convergence to a stable optimum cannot be established solely from the available training history. The observed behaviour nevertheless provides evidence of progressive learning and supports the feasibility of applying PPO to sequential resource block allocation in a simulated 5G wireless communication environment.

6.2 Spectral Efficiency Performance Comparison

The comparative spectral efficiency performance of the proposed PPO-based resource allocation framework and the conventional scheduling algorithms is presented in Figure 6. The results demonstrate substantial differences in bandwidth utilization among the evaluated strategies. The PPO-based framework achieved an average spectral efficiency of approximately 7.8 bits/s/Hz, while the Round Robin and Random Allocation approaches achieved approximately 3.9- and 4.2-bits/s/Hz, respectively. The Max-SINR scheduling algorithm exhibited a spectral efficiency of approximately 7.7 bits/s/Hz, demonstrating performance close to that of the proposed learning-based approach.

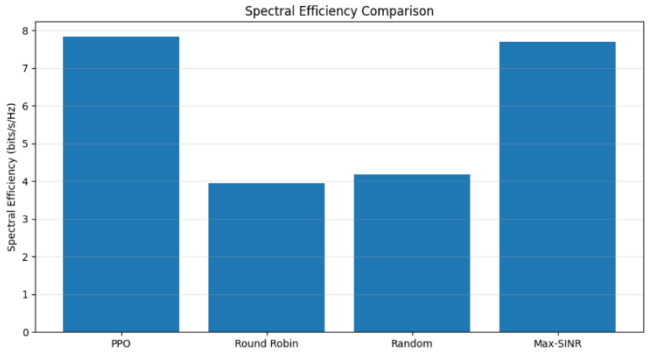


Figure 7: Spectral efficiency comparison

The percentage improvement calculations obtained from the simulation indicate that the PPO framework achieved a 98.51% improvement in spectral efficiency compared with Round Robin scheduling and an 87.67% improvement compared with Random Allocation. These improvements indicate that adaptive resource allocation can utilize available wireless communication resources more effectively than scheduling mechanisms that do not explicitly optimize channel-dependent transmission opportunities. The comparatively low spectral efficiency obtained using Round Robin scheduling can be attributed to its predetermined resource distribution mechanism, which does not prioritize users according to instantaneous channel quality. Similarly, Random Allocation does not systematically exploit favourable propagation conditions, resulting in reduced aggregate spectral utilization.

The comparison with Max-SINR scheduling is particularly important because this benchmark explicitly selects users experiencing favourable channel conditions. The PPO framework achieved a comparatively modest spectral efficiency improvement of 1.96% over Max-SINR, indicating that the learning-based approach performed slightly better under the evaluated simulation conditions. This result suggests that the PPO agent learned a resource allocation policy capable of exploiting channel quality information while considering additional reward components associated with fairness and interference. Nevertheless, the small performance difference also demonstrates the effectiveness of conventional channel-aware scheduling in the considered single-cell environment. Without statistical confidence intervals or additional independent training runs, the observed improvement over Max-SINR should be interpreted as a measured simulation

outcome rather than conclusive evidence of general superiority.

The results establish that the proposed PPO framework is capable of achieving high spectral efficiency under the selected network configuration. The substantial differences relative to Round Robin and Random Allocation highlight the importance of channel-aware scheduling, while the close comparison with Max-SINR demonstrates that conventional optimization principles remain highly competitive. These findings suggest that the principal value of learning-based resource management may become more apparent in extended network environments involving additional constraints, heterogeneous service requirements, or more complex allocation decisions.

6.3 Network Throughput Performance Analysis

The network throughput comparison illustrated in Figure 7 provides an additional assessment of the communication capacity achieved by the evaluated resource allocation algorithms. The PPO-based framework achieved an aggregate network throughput of approximately 16.9 Mbps, representing the highest throughput among the four investigated scheduling approaches. The Max-SINR algorithm achieved approximately 16.6 Mbps, while Random Allocation and Round Robin scheduling produced approximately 9.0 Mbps and 8.5 Mbps, respectively. These results are consistent with the spectral efficiency comparison and demonstrate that the proposed learning-based scheduling strategy achieved higher aggregate transmission rates under the considered simulation conditions.

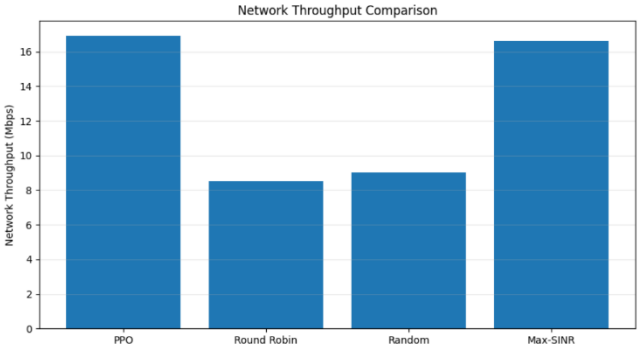


Figure 8: Network throughput comparison

The relatively high throughput obtained using PPO indicates that the trained agent learned to allocate resource blocks in a manner that effectively utilized favourable channel conditions. By incorporating channel quality observations and accumulated user throughput into the decision-making process, the PPO framework was able to select scheduling actions associated with higher aggregate achievable rates. In contrast, Round Robin scheduling distributed resources according to a fixed sequence without considering instantaneous channel variations, thereby limiting its ability to exploit strong communication channels. Random Allocation exhibited slightly higher throughput than Round Robin in the presented results, although its decisions were not guided by a systematic optimization objective.

The Max-SINR scheduling algorithm achieved throughput close to that of PPO because its channel-aware selection mechanism directly prioritizes users with strong instantaneous signal quality. The comparatively small difference between these methods indicates that the proposed PPO framework and the Max-SINR benchmark achieved similar aggregate communication capacity in the simulated environment. Furthermore, the correspondence between the throughput and spectral efficiency results is expected because both metrics were calculated using the same fixed total communication bandwidth. Consequently, the throughput comparison supports the spectral efficiency findings but does not constitute an independent measure of a different optimization outcome.

Overall, the results indicate that PPO-based intelligent scheduling can achieve substantial throughput improvements over non-channel-aware resource allocation strategies. However, the findings also demonstrate that improvements in aggregate throughput must be interpreted alongside user fairness, since concentrating resources on favourable channels may increase total network capacity while limiting communication opportunities for users experiencing weaker propagation conditions.

6.4 User Fairness Performance Evaluation

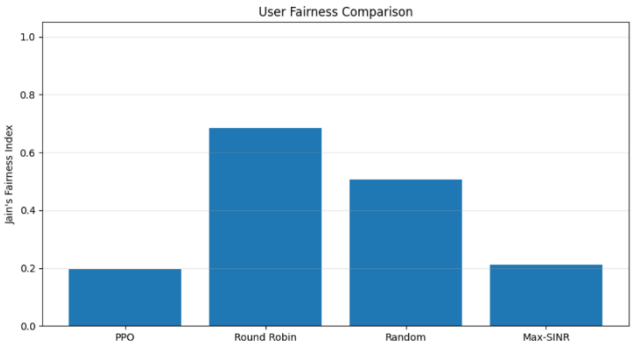


Figure 9: Jain's fairness index comparison

The user fairness comparison presented in Figure 8 reveals an important limitation of the proposed PPO-based resource allocation framework. Jain's fairness index was employed to quantify the degree of equality in achieved transmission rates among the connected users. The results indicate that

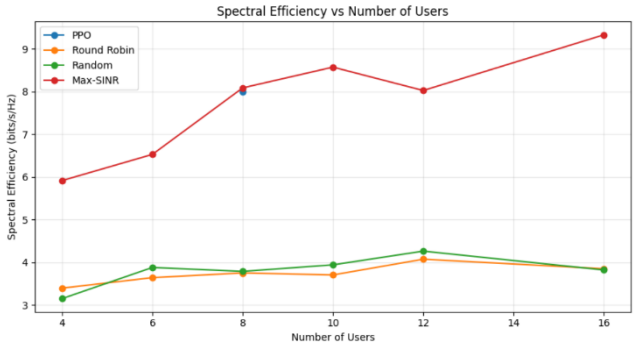
Round Robin scheduling achieved the highest fairness index, approximately 0.68, followed by Random Allocation at approximately 0.51. In contrast, the PPO-based framework achieved a fairness index of approximately 0.20, while Max-SINR scheduling produced a similarly low value of approximately 0.21.

The comparatively high fairness obtained using Round Robin scheduling can be attributed to its cyclic resource assignment mechanism, which distributes scheduling opportunities among users without systematically favouring those experiencing strong wireless channels. Although this approach produced lower aggregate spectral efficiency and throughput, it offered a more balanced distribution of transmission opportunities across the simulated network. Random Allocation achieved an intermediate fairness level because its resource assignments were distributed probabilistically among users, although the resulting allocation was not necessarily optimal in terms of either throughput or fairness.

The relatively low fairness indices obtained using PPO and Max-SINR indicate that both approaches concentrated a substantial proportion of achievable network throughput among a limited subset of users. In the case of Max-SINR scheduling, this behaviour follows directly from its preference for users with favourable instantaneous channel conditions. For PPO, the result suggests that the implemented reward formulation did not provide sufficient incentive to achieve equitable throughput distribution, despite explicitly incorporating a fairness-related component. The spectral efficiency reward may have exerted a stronger influence on the learned scheduling policy than the fairness reward, particularly when allocating resources to users with favourable propagation conditions produced substantial increases in achievable data rate.

This observation highlights an important trade-off between maximizing spectral efficiency and maintaining fairness in multi-user wireless communication systems. Although the proposed PPO framework achieved the highest spectral efficiency and network throughput among the evaluated algorithms, it did not demonstrate superior fairness performance. Therefore, the present results do not support a claim that the proposed model simultaneously optimizes both spectral efficiency and equitable resource distribution. Instead, they reveal an opportunity for improving the reward function through adaptive weighting, minimum user throughput constraints, fairness-aware scheduling penalties, or alternative multi-objective reinforcement learning techniques. Addressing this limitation could enhance the suitability of the proposed framework for heterogeneous 5G applications where reliable and equitable access to communication resources is essential.

6.5 Influence of User Density on Spectral Efficiency




```
SPECTRAL EFFICIENCY IMPROVEMENT
PPO vs Round Robin: 98.51%
PPO vs Random: 87.67%
PPO vs Max-SINR: 1.96%

All results saved successfully.
Results directory: /content/5G_DRL_Results
ZIP archive: /content/5G_DRL_Research_Results.zip
```

Figure 10: Spectral efficiency versus number of users

The relationship between the number of connected users and the achievable spectral efficiency of the evaluated scheduling algorithms is illustrated in Figure 9. The simulation considered network configurations containing between four and sixteen users, enabling an examination of how increasing user density influences the effectiveness of conventional resource allocation strategies. The results demonstrate that Max-SINR scheduling consistently achieved substantially higher spectral efficiency than Round Robin and Random Allocation across the evaluated user configurations. At four users, Max-SINR achieved approximately 5.9 bits/s/Hz, increasing to approximately 8.1 bits/s/Hz at eight users and reaching approximately 9.3 bits/s/Hz at sixteen users.

The increasing spectral efficiency observed for Max-SINR scheduling can be explained by the presence of multi-user diversity. As the number of connected users increases, the probability of identifying at least one user with favourable channel conditions on a particular resource block also increases. Since Max-SINR scheduling selects the user experiencing the strongest instantaneous signal quality, it can exploit these favourable channel opportunities to improve aggregate spectral utilization. The observed trend therefore reflects the benefits of channel-aware scheduling under increasing user populations, although the performance

is also influenced by the randomly generated user locations and fading conditions.

Round Robin and Random Allocation exhibited considerably lower spectral efficiency across the evaluated user configurations. Their results fluctuated within a relatively narrow range of approximately 3.2 to 4.3 bits/s/Hz, indicating that increasing the number of connected users did not consistently translate into improved spectral utilization for these allocation strategies. This behaviour is expected because neither method systematically selects the user with the most favourable instantaneous communication channel. The relatively small fluctuations between different user configurations can be associated with variations in the generated channel realizations and the limited number of evaluation episodes.

The PPO framework was evaluated only for the eight-user configuration, where it achieved spectral efficiency close to 8.0 bits/s/Hz. This restriction arises because the implemented neural network was trained using a fixed observation and action structure corresponding to eight users. Consequently, the current scalability experiment does not establish the performance of PPO under varying user densities. The single PPO result provides a reference point for comparison at the original network size, but it cannot be used to infer that the proposed learning-based algorithm would maintain its advantage as the number of connected users increases. A more comprehensive scalability investigation would require separately trained PPO agents for different user configurations or the development of a variable-dimensional neural network architecture.

The findings therefore demonstrate the effectiveness of Max-SINR scheduling in exploiting multi-user diversity and identify the need for additional experiments to establish the scalability of the proposed DRL framework. Future investigations should consider variable-size state representations, multi-agent reinforcement learning, or attention-based policy architectures to support intelligent resource allocation across different network densities.

6.6 Overall Comparative Performance and Discussion

The comparative analysis of the simulation results demonstrates that the proposed PPO-based resource allocation framework achieved the highest aggregate spectral efficiency and network throughput among the four investigated scheduling strategies. The learning-based approach exhibited substantial improvements over Round Robin and Random Allocation, while its performance remained close to that of Max-SINR scheduling. These findings indicate that deep reinforcement learning can develop effective channel-aware scheduling policies through repeated interactions with a stochastic wireless environment. The observed improvement in training reward further supports the ability of the PPO agent to learn resource allocation decisions associated with higher cumulative rewards.

A summary of the principal performance indicators is presented in Table 1. The numerical values are approximate readings from the supplied graphs, while the percentage improvements are taken from the simulation output.

Table 1. Comparative performance of resource allocation algorithms in the simulated 5G wireless network

Resource allocation algorithm	Spectral efficiency (bits/s/Hz)	Network throughput (Mbps)	Jain's fairness index
Proposed PPO	≈ 7.8	≈ 16.9	≈ 0.20
Round Robin	≈ 3.9	≈ 8.5	≈ 0.68
Random Allocation	≈ 4.2	≈ 9.0	≈ 0.51
Max-SINR	≈ 7.7	≈ 16.6	≈ 0.21

The results also reveal that improved aggregate spectral efficiency does not necessarily correspond to improved user fairness. While PPO and Max-SINR achieved high throughput, both exhibited relatively low fairness indices, indicating unequal distribution of achieved data rates among connected users. This finding is particularly significant for practical 5G communication networks, where different applications and users may have distinct minimum service requirements. The results therefore highlight the need for more carefully balanced reward formulations that incorporate explicit fairness objectives or minimum service guarantees.

Another important observation concerns the relatively small performance difference between PPO and Max-SINR scheduling. Although the proposed PPO framework achieved a 1.96% spectral efficiency improvement in the reported experiment, the available results do not establish whether this advantage would remain consistent across multiple independent training runs, different channel distributions, or more complex network configurations. Additional experiments involving repeated training with different random seeds, statistical confidence intervals, and sensitivity analysis would be necessary to evaluate the robustness of the observed improvement. Furthermore, the current model considers a simplified single-cell communication environment with equal transmission power

across resource blocks, and the learned agent performs resource block scheduling rather than joint power and bandwidth optimization.

Despite these limitations, the simulation establishes a useful computational foundation for investigating intelligent resource allocation in 5G wireless communication networks. The integration of a stochastic wireless environment, PPO-based decision-making, reward-driven optimization, and comparative scheduling evaluation provides a reproducible approach for examining the relationship between reinforcement learning and spectrum utilization. The findings suggest that learning-based scheduling can be competitive with conventional channel-aware resource management and substantially outperform allocation mechanisms that do not exploit instantaneous channel quality. Further development of the framework should prioritize fairness-aware reward design, adaptive power allocation, variable user populations, and evaluation under more realistic 5G propagation and interference conditions.

VII. CONCLUSION AND FUTURE WORK

The present research investigated the development and performance evaluation of a Deep Reinforcement Learning-Based Intelligent Resource Allocation framework for enhancing spectral efficiency in fifth-generation (5G) wireless communication networks. The study addressed the challenges associated with efficient radio resource utilization in multi-user wireless environments characterized by varying channel conditions, interference, and competing communication demands. A Proximal Policy Optimization (PPO)-based intelligent scheduling mechanism was developed using Python and evaluated within a simulated 5G downlink communication environment incorporating Rayleigh fading, distance-dependent path loss, and interference-aware channel modelling. The proposed framework employed a reward-driven learning strategy to identify adaptive resource block allocation policies by considering spectral efficiency, user fairness, and interference characteristics. Its effectiveness was assessed through comparative evaluation against conventional Round Robin, Random Allocation, and Maximum Signal-to-Interference-plus-Noise Ratio (Max-SINR) scheduling algorithms.

The simulation results demonstrated that the proposed PPO-based framework achieved improved aggregate spectral utilization and network throughput compared with the conventional scheduling approaches under the investigated network configuration. The developed model achieved approximately 7.8 bits/s/Hz spectral efficiency and 16.9 Mbps aggregate network throughput, representing improvements of 98.51% and 87.67% in spectral efficiency compared with Round Robin and Random Allocation, respectively. Furthermore, the proposed approach achieved a 1.96% improvement over Max-SINR scheduling, indicating that the learned allocation policy performed competitively with an established channel-aware resource management technique. The training reward convergence characteristics demonstrated a progressive improvement in the learning agent's decision-making performance, supporting the feasibility of PPO for adaptive resource allocation in stochastic wireless environments. These findings establish the potential of deep reinforcement learning to improve resource utilization through sequential, experience-driven decision-making rather than relying exclusively on predetermined scheduling rules.

An important outcome of the investigation was the identification of a significant trade-off between aggregate spectral efficiency and equitable resource distribution. Although the proposed PPO framework achieved the highest spectral efficiency among the evaluated algorithms, its Jain's fairness index remained approximately 0.20, considerably below the value achieved by Round Robin scheduling. This observation indicates that the adopted reward formulation prioritized aggregate communication performance more strongly than equitable throughput distribution. Consequently, the present findings emphasize that maximizing spectral efficiency alone is insufficient for ensuring balanced service delivery across heterogeneous wireless users. The investigation also demonstrated that conventional Max-SINR scheduling remains highly competitive in simplified single-cell environments, suggesting that the advantages of learning-based resource management should be examined under more complex communication conditions and additional operational constraints.

The principal contribution of this research lies in the integration of stochastic wireless channel modelling, PPO-based sequential resource block scheduling, multi-objective reward formulation, and comparative performance evaluation within a unified Python-based simulation framework. The developed methodology provides a reproducible computational foundation for examining the relationship between intelligent scheduling decisions and achievable spectral utilization in multi-user 5G communication systems. While the present implementation is limited to a simplified single-cell downlink environment with fixed transmission power distribution and a predefined number of users, the obtained results demonstrate the feasibility of applying deep reinforcement learning to

adaptive wireless resource management. Overall, the study establishes that PPO-based intelligent scheduling can provide substantial performance improvements over non-channel-aware allocation strategies while identifying fairness, scalability, and robustness as essential considerations for further development. The findings contribute to the broader investigation of artificial intelligence-enabled radio resource management and provide a foundation for designing more adaptive and efficient communication systems.

Future research will focus on extending the proposed deep reinforcement learning framework to address the limitations identified during the present investigation and improve its suitability for more realistic 5G wireless communication environments. A primary direction involves developing an advanced fairness-aware reward mechanism capable of achieving a more balanced relationship between spectral efficiency, network throughput, and equitable resource distribution. Adaptive reward weighting, minimum user throughput constraints, and multi-objective reinforcement learning techniques may be investigated to prevent excessive resource concentration among users experiencing favourable channel conditions. Further improvements may be achieved by incorporating joint resource block and transmission power optimization, enabling the intelligent agent to make coordinated decisions concerning bandwidth utilization, interference management, and energy consumption.

Another important extension involves evaluating the proposed framework under varying user densities, heterogeneous traffic demands, and dynamic mobility conditions. Since the present PPO implementation employs a fixed observation and action structure, future investigations may explore graph neural networks, attention-based reinforcement learning architectures, and multi-agent learning techniques capable of supporting variable numbers of connected users. The incorporation of realistic 5G channel models, multi-cell interference, user mobility, and service-specific quality-of-service requirements would further strengthen the practical relevance of the proposed methodology. Comparative investigations involving alternative deep reinforcement learning algorithms, including Double Deep Q-Networks, Soft Actor-Critic, and advanced actor-critic architectures, could provide additional insights into training stability, computational efficiency, and policy generalization.

Future work should also include repeated experiments using independent random seeds, statistical confidence intervals, sensitivity analysis, and evaluation under previously unseen network conditions to establish the robustness of the learned allocation policy. The integration of energy efficiency objectives, latency constraints, and intelligent network slicing could extend the framework toward comprehensive radio resource management for heterogeneous 5G services. Additionally, validation using established wireless network simulation platforms and software-defined radio testbeds would help assess the feasibility of transferring simulation-based learning policies to more realistic communication environments. Ultimately, these developments could support the evolution of autonomous, AI-enabled radio access networks and provide useful research directions for emerging sixth-generation (6G) communication systems, where intelligent resource orchestration, adaptive spectrum utilization, and sustainable network operation are expected to play increasingly important roles.

REFERENCES

[1] J. G. Andrews et al., “What will 5G be?,” IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] M. Agiwal, A. Roy, and N. Saxena, “Next generation 5G wireless networks: A comprehensive survey,” IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1617–1655, 2016.

[3] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” IEEE Communications Magazine, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[4] A. Gupta and R. K. Jha, “A survey of 5G network: Architecture and emerging technologies,” IEEE Access, vol. 3, pp. 1206–1232, 2015.

[5] N. Panwar, S. Sharma, and A. K. Singh, “A survey on 5G: The next generation of mobile communication,” Physical Communication, vol. 18, pp. 64–84, Mar. 2016.

[6] M. Shafi et al., “5G: A tutorial overview of standards, trials, challenges, deployment, and practice,” IEEE Journal on Selected Areas in Communications, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[7] T. S. Rappaport et al., “Millimeter wave mobile communications for 5G cellular: It will work!,” IEEE Access, vol. 1, pp. 335–349, 2013.

[8] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” IEEE Communications Magazine, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[9] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Transactions on Wireless Communications, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[10] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," IEEE Transactions on Communications, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[11] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," Foundations and Trends in Signal Processing, vol. 11, no. 3–4, pp. 154–655, 2017.

[12] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," IEEE Transactions on Communications, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[13] D. Gesbert, S. G. Kiani, A. Gjendemsjø, and G. E. Øien, "Adaptation, coordination, and distributed resource allocation in interference-limited wireless networks," Proceedings of the IEEE, vol. 95, no. 12, pp. 2393–2409, Dec. 2007.

[14] H. Holma and A. Toskala, LTE for UMTS: Evolution to LTE-Advanced, 2nd ed. Chichester, U.K.: Wiley, 2011.

[15] J. Huang, V. G. Subramanian, R. Agrawal, and R. A. Berry, "Downlink scheduling and resource allocation for OFDM systems," IEEE Transactions on Wireless Communications, vol. 8, no. 1, pp. 288–296, Jan. 2009.

[16] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks—Part I: Theoretical framework," IEEE Transactions on Wireless Communications, vol. 4, no. 2, pp. 614–624, Mar. 2005.

[17] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks—Part II: Algorithm development," IEEE Transactions on Wireless Communications, vol. 4, no. 2, pp. 625–634, Mar. 2005.

[18] Z. Han, Z. Ji, and K. J. R. Liu, "Fair multiuser channel allocation for OFDMA networks using Nash bargaining solutions and coalitions," IEEE Transactions on Communications, vol. 53, no. 8, pp. 1366–1376, Aug. 2005.

[19] H. Zhang, C. Jiang, N. C. Beaulieu, X. Chu, X. Wang, and T. Q. S. Quek, "Resource allocation for cognitive small cell networks: A cooperative bargaining game theoretic approach," IEEE Transactions on Wireless Communications, vol. 14, no. 6, pp. 3481–3493, Jun. 2015.

[20] M. Chiang, P. Hande, T. Lan, and C. W. Tan, "Power control in wireless cellular networks," Foundations and Trends in Networking, vol. 2, no. 4, pp. 381–533, 2008.

[21] N. C. Luong et al., "Applications of deep reinforcement learning in communications and networking: A survey," IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3133–3174, 2019.

[22] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[23] V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[24] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in Proc. AAAI Conf. Artificial Intelligence, vol. 30, no. 1, pp. 2094–2100, 2016.

[25] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," in Proc. 4th International Conference on Learning Representations (ICLR), 2016.

[26] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.

[27] J. Schulman, S. Levine, P. Moritz, M. I. Jordan, and P. Abbeel, "Trust region policy optimization," in Proc. 32nd International Conference on Machine Learning, pp. 1889–1897, 2015.

[28] J. Schulman, P. Moritz, S. Levine, M. I. Jordan, and P. Abbeel, "High-dimensional continuous control using generalized advantage estimation," in Proc. 4th International Conference on Learning Representations (ICLR), 2016.

[29] J. Achiam, D. Held, A. Tamar, and P. Abbeel, "Constrained policy optimization," in Proc. 34th International Conference on Machine Learning, pp. 22–31, 2017.

[30] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in Proc. 35th International Conference on Machine Learning, pp. 1861–1870, 2018.

[31] V. Mnih et al., "Asynchronous methods for deep reinforcement learning," in Proc. 33rd International Conference on Machine Learning, pp. 1928–1937, 2016.

[32] Z. Wang et al., "Dueling network architectures for deep reinforcement learning," in Proc. 33rd International Conference on Machine Learning, pp. 1995–2003, 2016.

[33] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," in Proc. 4th International Conference on Learning Representations (ICLR), 2016.

[34] Y. S. Nasir and D. Guo, "Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks," IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp. 2239–2250, Oct. 2019.

[35] F. Meng, P. Chen, L. Wu, and J. Cheng, "Power allocation in multi-user cellular networks: Deep reinforcement learning approaches," IEEE Transactions on Wireless Communications, vol. 19, no. 10, pp. 6255–6267, Oct. 2020.

[36] H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," IEEE Transactions on Vehicular Technology, vol. 68, no. 4, pp. 3163–3173, Apr. 2019.

[37] A. Alwarafy, M. Abdallah, B. S. Ciftler, A. Al-Fuqaha, and M. Hamdi, "The frontiers of deep reinforcement learning for resource management in future wireless HetNets: Techniques, challenges, and research directions," IEEE Open Journal of the Communications Society, vol. 3, pp. 322–365, 2022.

[38] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for interference management," IEEE Transactions on Signal Processing, vol. 66, no. 20, pp. 5438–5453, Oct. 2018.

[39] Q. Huang, X. Xie, and M. Cheriet, "Reinforcement learning-based hybrid spectrum resource allocation scheme for the high load of URLLC services," EURASIP Journal on Wireless Communications and Networking, vol. 2020, Art. no. 250, Dec. 2020.

[40] D. Wang, H. Qin, B. Song, K. Xu, X. Du, and M. Guizani, "Joint resource allocation and power control for D2D communication with deep reinforcement learning in MCC," Physical Communication, vol. 45, Art. no. 101262, Apr. 2021.

[41] R. Jain, D.-M. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Digital Equipment Corporation, Hudson, MA, USA, Research Rep. DEC-TR-301, Sep. 1984.

[42] D. P. Bertsekas, Dynamic Programming and Optimal Control, vol. 1, 4th ed. Belmont, MA, USA: Athena Scientific, 2017.

[43] M. L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming. New York, NY, USA: Wiley, 1994.

[44] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge, U.K.: Cambridge University Press, 2005.

[45] A. Goldsmith, Wireless Communications. Cambridge, U.K.: Cambridge University Press, 2005.

[46] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge University Press, 2004.

CORRESPONDING AUTHOR: Miss. DEEPIKA PANDA