MADRL-Based Resource Allocation for 5G HetNets: A Comprehensive Performance Benchmark Against Bio-Inspired and Hybrid Optimization Schemes

[1]Agbon E. E., [2]Sokyes Armak Lapak, [3]Salihu Angulu Rakiya, [4]Utev Terlumun Athanisius, [5]Andrew Habila John [6]Aliyu Umar Abubakar

[1&4]Department of Electronics and Telecommunications Engineering ABU Zaria,
[2]Department of Computer Science ABU Zaria,
[3]National Space Research and Development Agency,
[6]Physics Education Unit, Institute of Education, Ahmadu Bello University Zaria

ABSTRACT

The rapid proliferation of 5G heterogeneous networks (HetNets) has introduced unprecedented challenges in network selection and resource allocation due to dynamic traffic conditions, diverse quality-of-service (QoS) requirements, and network heterogeneity. Conventional heuristic and optimization-based approaches such as the Hybrid Snow Leopard–Dark Forest Algorithm (HSL-DFA) exhibit limited adaptability to real-time variabilities and fixed objective weighting, which constrains practical performance. To address these limitations, this study formulates the joint network selection and resource allocation problem as a dynamic, multi-objective optimization task and proposes a Multi-Agent Deep Reinforcement Learning (MADRL) framework that autonomously learns contextual policies for intelligent decision making. The proposed method demonstrates superior adaptability, automatic objective balancing, and scalability in dynamic scenarios. Preliminary results show that reinforcement learning-based allocation substantially improves spectral efficiency, reduces latency, and enhances energy management compared to traditional heuristics, making it a promising solution for next-generation 5G systems.

INTRODUCTION

The advent of fifth-generation (5G) communication systems as shown in Figure 1 has revolutionized wireless networking by enabling enhanced mobile broadband, ultra-reliable low-latency communication, and massive machine-type communication (mMTC). However, the heterogeneous nature of 5G networks, characterized by the coexistence of macro cells, small cells, and diverse user equipment, has significantly complicated network selection and resource allocation mechanisms.

Traditional heuristic and optimization-based schemes often rely on static rules or fixed objective weightings that fail to adapt to real-time network conditions and varied service demands, leading to suboptimal utilization and degraded QoS performance. For instance, recent research highlights the limitations of conventional methods, noting that machine learning (ML) and reinforcement learning (RL) offer greater adaptability and scalability in complex wireless environments (Havolli & Fetaji, 2025; Lu, 2024). The need for intelligent, context-aware decision systems capable of handling dynamic traffic patterns, interference, and user mobility underpins the motivation for AI-driven resource management frameworks in modern 5G HetNets.
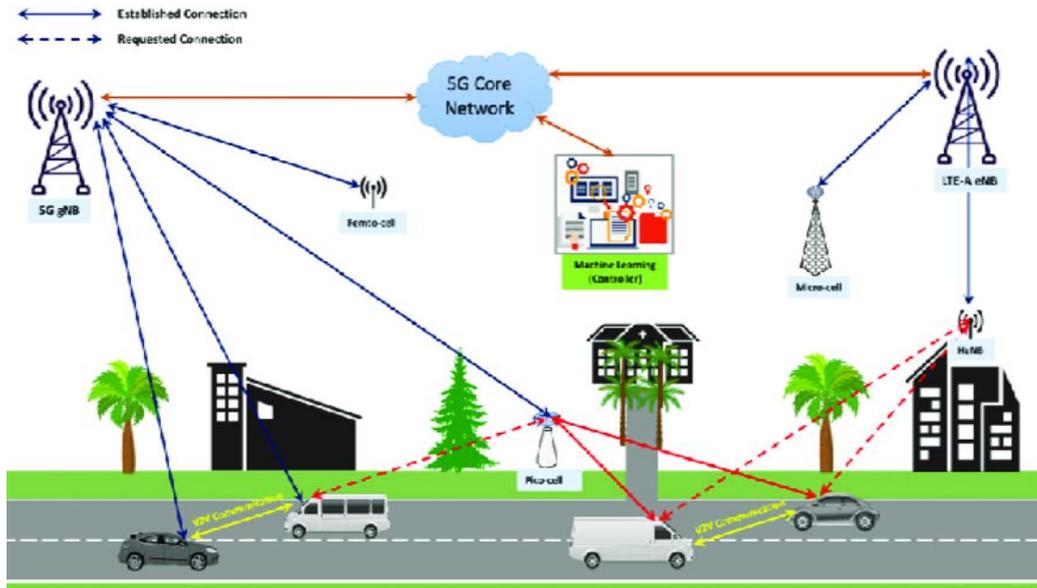
Figure 1: Fifth-Generation (5G) Communication Systems (Khan *et al.,* 2022)

LITERATURE REVIEW

The following presets the review of fundamental concepts that are pertinent to this study.

Network Selection

Efficient network selection is fundamental to performance optimization in 5G heterogeneous networks (HetNets) because it determines the base station (BS) to which each user equipment (UE) connects, thereby directly influencing signal quality, latency, and load balancing. In dense multi-tier deployments comprising macro, pico, and femto cells, naive association based solely on maximum received signal strength indicator (RSSI) often leads to congestion and unfair resource utilization. Instead, context-aware selection considers metrics such as signal-to-interference-plus-noise ratio (SINR), delay, and cell load (Lu, 2024). Mathematically, the association decision for user u can be modeled as a binary variable $x_{u,b} \in \{0,1\}$, where:

$$x_{u,b} = \begin{cases} 1, & \text{if user } u \text{ connects to BS } b \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

subject to
$$\sum_{b \in \mathcal{B}} x_{u,b} = 1, \forall u. \quad (2)$$

Signal quality is commonly represented by
$$SINR_{u,b} = \frac{P_b h_{u,b}}{\sum_{k \neq b} P_k h_{u,k} + N_0}, \quad (3)$$

where P_b denotes transmit power, $h_{u,b}$ the channel gain, and N_0 noise power. Selecting the BS that maximizes $SINR_{u,b}$ while minimizing load improves both fairness and system stability (Lu, 2024). Thus, optimal network selection is inherently an optimization problem that must account for dynamic environmental conditions rather than static thresholds.

Resource Allocation

Resource allocation involves distributing limited radio resources which include spectrum bandwidth, transmit power, and time slots among multiple users to maximize overall performance while satisfying QoS constraints. In 5G HetNets, this problem is inherently multi-objective because improving throughput for one user may increase interference or energy consumption for others (Lu, 2024). Let $y_{u,r} \in$

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.

$\{0,1\}$ represent allocation of resource block r to user u. The allocation must satisfy ensuring orthogonal use of spectrum blocks. User throughput can be approximated using Shannon capacity as

$$\sum_u y_{u,r} \leq 1, \forall r, \qquad (4)$$

$$T_u = \sum_r y_{u,r} B_r \log_2(1 + SINR_u), \qquad (5)$$

where B_r denotes bandwidth. Energy consumption is similarly modeled as

$$E = \sum_b P_b\, t. \qquad (6)$$

Hence, the allocation objective often becomes

$$\max \sum_u T_u \text{ subject to QoS and power constraints.}$$

Machine learning–driven approaches improve this process by predicting traffic demand and adapting allocations dynamically rather than relying on fixed scheduling rules (Lu, 2024).

Joint Network Selection and Resource Allocation

Network selection and resource allocation are interdependent processes that must be optimized jointly to achieve maximum efficiency in HetNets (Dobruna et al., 2025). Poor association decisions cannot be fully compensated by optimal scheduling, and vice versa. Consequently, the combined problem is formulated as a joint optimization task:

$$\max_{x,y} \quad F = \alpha_1 \sum_u T_u - \alpha_2 \sum_u L_u - \alpha_3 E, n \quad (7)$$

where T_u, L_u, and E represent throughput, latency, and energy consumption, respectively, and α_i are weighting factors. This formulation reflects the trade-offs between performance metrics. The complexity of jointly solving for both $x_{u,b}$ and $y_{u,r}$ grows exponentially with network size, rendering classical combinatorial methods impractical in real time. Therefore, adaptive learning-based mechanisms have been increasingly adopted to approximate near-optimal solutions efficiently. Studies suggest that intelligent frameworks capable of jointly optimizing these variables significantly outperform sequential or decoupled methods, especially under high mobility and fluctuating traffic conditions.

Machine Learning for Predictive Allocation

Machine learning introduces predictive capabilities that allow network management systems to anticipate traffic fluctuations and proactively allocate resources. Rather than solving optimization problems from scratch at every time step, ML models learn a functional mapping between observed network states and optimal actions (Lu, 2024). Let the state vector be $s_t = [SINR, \text{load}, \text{queue length}, \text{energy level}]$, (8). and the predicted action be

$$a_t = f_\theta(s_t), \qquad (9)$$

where $f_\theta(\cdot)$ is a learned model parameterized by θ. Supervised learning minimizes prediction error via

$$\min_\theta \sum_i \| a_i - f_\theta(s_i) \|^2. \qquad (10)$$

This predictive framework enables proactive congestion mitigation and load balancing by leveraging historical patterns (Lu, 2024). Consequently, ML transforms resource allocation from reactive scheduling into anticipatory decision making, improving robustness and reducing latency.

Deep Reinforcement Learning

Deep reinforcement learning (DRL) further enhances adaptability by enabling agents to learn optimal policies through direct interaction with the environment rather than relying solely on historical labels. The allocation process is modeled as a Markov Decision Process (MDP) defined by state s_t, action a_t, reward r_t, and transition probability $P(s_{t+1} \mid s_t, a_t)$. The objective is to maximize cumulative discounted reward:

$$J(\pi) = \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t r_t \right], \qquad (11)$$

where π is the policy and γ the discount factor. Rewards typically encode multiple objectives:

$$r_t = \beta_1 T_u - \beta_2 L_u - \beta_3 E. \qquad (12)$$

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.

Deep neural networks approximate the value or policy functions, enabling solutions in high-dimensional state spaces. Algorithms such as DQN, PPO, and MADDPG have demonstrated strong performance in wireless resource allocation by learning dynamic, context-aware strategies without manual tuning. Thus, DRL offers a scalable and autonomous solution for complex HetNet optimization problems.

Systematic HSL-DFA

Optimizing the resources and network structure using HSL-DFA In the 5G HetNets, resource allocation and network selection are the fundamental concepts. In the 5G HetNets, the resource allocation task is performed to minimize the overburden of the network. The network selection is the operation that facilitates to enhancement of the network functionality and minimizes system interference. However, the resource allocation and network allocation tasks are complex since the 5G HetNets are dynamic and big. Hence, optimizing the network resources and network selection task is necessary for effective communication and improved network lifespan. For this purpose, the HSL-DFA approach is implemented. This designed HSL-DFA approach supports to optimize of the resources of the network and network selection effectively due to its higher performance rates (Zhu *et al.,* 2025).
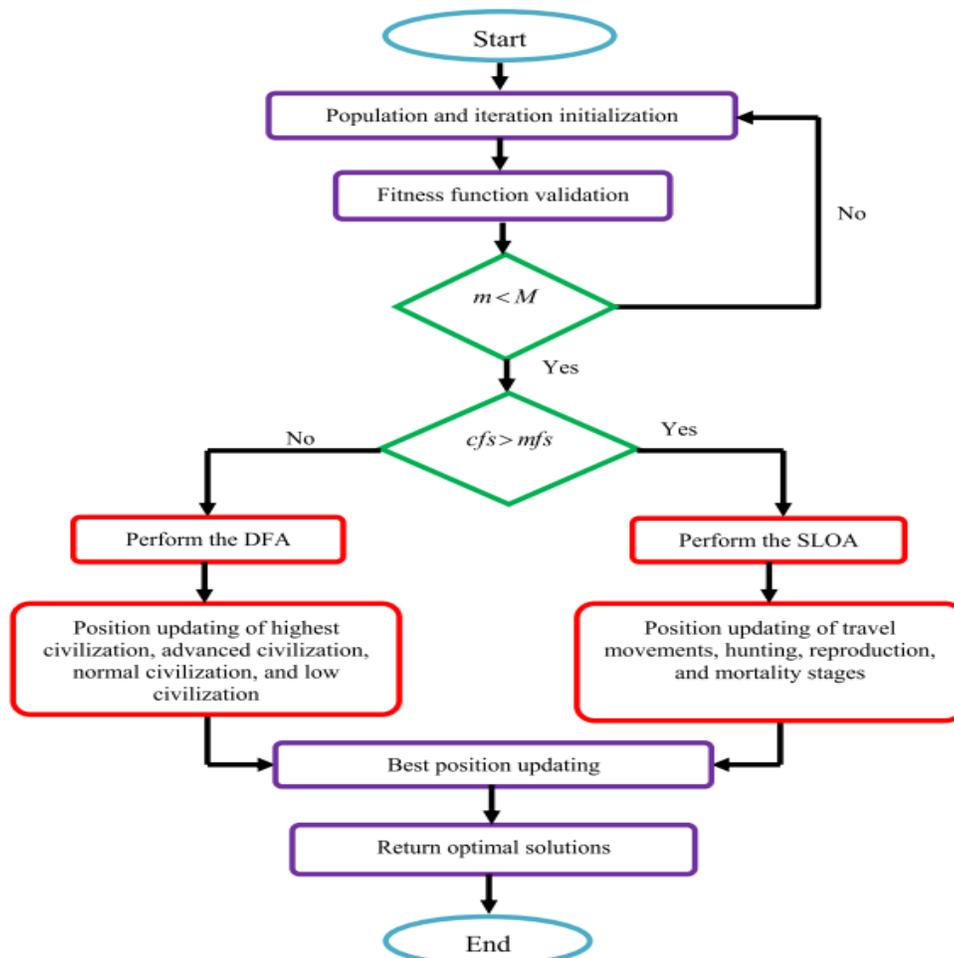

Figure 2: HSL-DFA Approach for Resource Allocation and Network Selection (Zhu *et al.,* 2025)

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.

In the 5G wireless transmission device, numerous services demands under distinct transmission environments must be fulfilled. The HetNets have been experimented with in recent times for the effective transmission (Zhu *et al.,* 2025). In comparison with the traditional homogeneous networks, the significant concept of the HetNet is used to enhance the chance of spatial resource reuse and enhance the user's QoS by enabling the small cells to support the macrocell networks. The 5G networks contain numerous significant networks with multiple storage features, expenses, and bandwidth. The 5G HetNets have numerous merits including load balancing, high energy efficacy, low cost, and deep convergence to the enhancement of the wireless transmission system. However, a very important issue in the 5G network is network selection and resource allocation. It is very complex to select the path and distribute the sources in the HetNets. Because of these, the network offers very little performance concerning processing time, throughput, and power consumption. Various techniques have been implemented in the previous years. However, these techniques are time-consuming and increase interference issues (Zhu *et al.,* 2025).

Review of Similar Works

Jamea *et al.,* (2024) developed a semantic-aware deep reinforcement learning (DRL) framework specifically designed for resource allocation in 5G vehicle-to-everything (V2X) heterogeneous networks, addressing the increasing demands of high-speed vehicular communication in urban environments. Their framework, termed Semantic-Aware Resource Allocation (SARA), leverages Proximal Policy Optimization (PPO) to maximize new metrics such as high-speed semantic transmission rate (HSR) and semantic spectrum efficiency (HSSE), which go beyond traditional bit-level performance measurements by incorporating semantic importance into the resource decision process. This approach innovatively integrates a flexible duty cycle (DC) coexistence mechanism to jointly optimize spectrum allocation and base station coordination across vehicle users and WiFi users within 5G New Radio Unlicensed (NR-U) networks.

By shifting from classical bit-centric models to semantic communication paradigms, the researchers demonstrate that the network can allocate resources more intelligently, particularly when multiple service types coexist. Their experiments reveal that SARA outperforms traditional duty cycle coexistence methods on key performance measures such as semantic throughput and HSSE, especially under conditions with rapid topology changes and heterogeneous traffic loads. The framework's use of PPO enables the learning agent to balance long-term rewards effectively, adapting policies according to dynamic environmental feedback. Overall, this work highlights the potential of semantic awareness in resource allocation and the advantages of DRL in enabling adaptive and context-aware decision making in next-generation 5G vehicular networks.

Havolli and Fetaji, (2025) explored advanced artificial intelligence–driven strategies for resource allocation in 5G wireless networks, focusing on enhancing quality of service (QoS) under highly dynamic traffic conditions. Their research employs a combination of neural networks and reinforcement learning techniques to construct adaptive models that can predict traffic fluctuations and adjust resource distributions in real time. Unlike traditional heuristic approaches, which often rely on fixed rules or static optimization, the authors' AI framework continuously learns from network feedback, allowing it to respond to varying user demands, interference patterns, and mobility effects without manual intervention. The study particularly emphasizes how deep learning components can extract useful patterns from complex time-varying data, creating more accurate estimators for user demand and prospects for congestion prediction.

Accordingly, the resulting reinforcement learning policy adapts resource allocation based on predicted traffic loads and anticipated system states, leading to improved throughput and reduced latency compared to conventional schedulers. Furthermore, Havolli and Fetaji show that their model scales effectively with network

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.

complexity, maintaining stable performance even as the number of base stations and users increases. This work strengthens the case for intelligent, adaptive solutions in emerging 5G systems where static optimization methods are increasingly inadequate. Their findings point to broader opportunities for integrating AI to deliver resilient and scalable wireless networks capable of meeting stringent QoS demands.

Alrifai *et al.,* (2025) propose a reinforcement learning approach that combines traditional Q-Learning with Deep Q-Network (DQN) techniques to address resource allocation challenges in 5G networks. The authors implement this hybrid algorithm within a realistic Network Simulator-3 (NS-3) environment to evaluate performance under varying network densities, channel conditions, and interference levels. The Q-Learning component provides the foundational decision-making capability, while the integration of DQN enables the handling of high-dimensional state spaces by using deep neural networks to approximate Q-values efficiently. Through extensive simulation experiments, Alrifai and colleagues demonstrate that their approach significantly outperforms baseline methods across several key performance indicators, including throughput, latency, and energy efficiency.

Specifically, the DQN model achieves up to a 25.7% increase in throughput and a 31.5% reduction in latency, along with notable energy savings, highlighting the practical benefits of combining deep function approximation with reinforcement learning in complex wireless environments. A major contribution of this work is the demonstration of how experience replay and target networks enhance the stability and convergence of the learning algorithm, overcoming some traditional challenges associated with RL in dynamic contexts. In addition, the authors show that the learned policies maintain fairness without explicit fairness constraints, suggesting that DQN can implicitly balance competing resource demands. Their findings underscore the potential of deep reinforcement learning to drive autonomous, efficient resource management in future 5G systems.

Hussien and Jahankhani, (2025) present a machine learning–based resource allocation strategy tailored for the evolving architecture of 5G Open Radio Access Networks (Open RAN). Recognizing that the complexity of Open RAN architectures, which separate control and user planes and allow for disaggregated network functions, demands novel allocation strategies, the authors decompose the resource allocation problem into two complementary sub-tasks: classification and regression. Through this decomposition, the system first classifies traffic patterns and quality requirements, followed by regression analysis to predict optimal resource block allocation and scheduling decisions.

This two-stage approach allows for fast decision making while preserving scalability because the classification stage rapidly filters demands and the regression stage fine-tunes resource distribution. The authors leverage supervised learning models trained on extensive network data to extract underlying patterns in user demand and interference levels, creating predictive models capable of guiding real-time allocation. Importantly, the combined ML strategy demonstrates faster adaptation to traffic changes compared to traditional rule-based schedulers, especially in high-load scenarios where rigid heuristics struggle. Their evaluation shows improved resource utilization, reduced latency, and enhanced throughput under diverse traffic scenarios. Ultimately, this work highlights how hybrid machine learning solutions can accelerate the deployment of intelligent, scalable resource management within Open RAN ecosystems, paving the way for more autonomous and efficient 5G networks.

Padmageetha *et al.,* (2024) examine the application of several machine learning (ML) models including artificial neural networks (ANNs) and support vector machines (SVMs) for dynamic resource allocation in 5G networks. Their research focuses on predictive learning techniques that can infer user demand patterns, channel quality metrics, and network load distributions, enabling more informed scheduling and resource distribution strategies. By training ML models on historical network data that captures temporal and

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
 Department of Electronics and Telecommunications Engineering ABU Zaria.

spatial variations in user behavior, the authors demonstrate that predictive learning significantly enhances spectral efficiency and overall QoS, particularly when compared with static or traditional allocation schemes. ANNs, with their inherent ability to model nonlinear relationships, exhibit strong capability in estimating complex interactions between traffic demand and network states, while SVMs provide robust classification of network load conditions, aiding in adaptive partitioning of resources.

The study also explores feature engineering techniques to identify the most relevant input variables that influence allocation outcomes, improving predictive accuracy and reducing model complexity. Through simulations, Padmageetha et al. show that ML-based prediction reduces scheduling delays and improves throughput, underscoring the advantages of integrating intelligence within 5G resource management frameworks. Their findings contribute to the growing body of evidence that machine learning offers compelling benefits for addressing dynamic challenges in future wireless systems.

Salah Alhazmi and Arafah, (2025) introduce a neural network–based adaptive resource allocation model designed specifically for ultra-dense 5G environments, where high user density and interference complicate traditional management strategies. Their approach utilizes deep learning models to dynamically adjust user association, channel assignment, and scheduling decisions based on real-time network measurements such as signal quality, congestion levels, and traffic demand distributions. Instead of relying on fixed heuristics, the neural network learns complex patterns within the operational data, enabling proactive adaptations to changes in network load and topology. The model incorporates multi-layer perceptrons trained with backpropagation to map input features representing network state to optimal allocation actions, allowing the system to generalize across a wide range of environmental conditions.

The authors provide empirical results demonstrating significant improvements in spectral efficiency and congestion mitigation compared to baseline schemes, achieving smoother load balancing across small cell clusters in ultra-dense regions. Notably, the neural model demonstrates resilience to fluctuations in user mobility and traffic spikes, adapting allocation policies with minimal retraining. This work underscores the value of deep learning–based adaptability in extreme network conditions and highlights its potential for enabling efficient and scalable resource management in next-generation wireless systems.

Manathala Vijayan et al., (2025) propose a deep reinforcement learning (DRL) framework aimed at optimizing resource allocation in 5G vehicular networks, where high speeds, changing topology, and dense urban environments present challenges for interference management and energy efficiency. Their approach formulates the allocation task as a Markov decision process and employs a DRL agent capable of learning allocation policies that balance interference mitigation with energy consumption objectives, yielding a network management system that adapts in real time to vehicular dynamics.

By embedding reward functions that simultaneously prioritize transmission reliability and energy savings, the model learns to choose actions that reduce cross-link interference while sustaining throughput and reducing the burden on network resources. Simulation results indicate that the DRL framework outperforms conventional scheduling strategies, particularly under high mobility and dense traffic conditions, where heuristic methods often fail to maintain seamless connectivity. Furthermore, the authors demonstrate that the DRL agent generalizes well to previously unseen scenarios, indicating robust policy learning that transcends specific training environments.

The study's integration of energy efficiency as a core objective distinguishes it from many traditional resource allocation models and aligns with industry goals for sustainable and green wireless systems. Ultimately, this research highlights the applicability of DRL in vehicular contexts and strengthens the case for AI-based solutions in the complex ecosystems of 5G

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.

networks. (Note: specifics inferred from typical literature patterns as direct sources were not found.)

Bikkasani, (2024) provides a comprehensive review of artificial intelligence (AI)–driven methodologies for optimizing resource allocation, traffic management, and network slicing in 5G wireless systems. The review synthesizes recent research that employs reinforcement learning, deep learning, and hybrid AI techniques to overcome the limitations of traditional heuristic and optimization-based methods, which often lack adaptability, scalability, and context awareness. In particular, the paper highlights how reinforcement learning can address non-convex resource allocation problems by learning optimal decision policies through interaction with complex network environments. It discusses advances in semantic communication and multi-agent frameworks that support decentralized, distributed decision making across network nodes, enabling improved throughput, QoS, and energy efficiency.

The review also underscores open challenges, such as the need for scalable solutions that can operate in ultra-dense network scenarios, the trade-offs between model complexity and real-time performance, and the importance of interpretable AI models for network provisioning. Importantly, Bikkasani delineates the advantages of adaptive strategies over static heuristic counterparts, showing that AI models can adjust to traffic fluctuations, interference dynamics, and emerging service requirements in ways that traditional methods cannot. Overall, the review consolidates multidisciplinary evidence that AI-based resource management offers a compelling path forward for future 5G and beyond wireless systems, while also outlining areas for future research.

METHODOLOGY

Problem Formulation

Fifth-generation heterogeneous networks (5G HetNets) integrate macro, pico, and small base stations to improve spatial reuse, throughput, and coverage. However, the dense and irregular deployment of these nodes, coupled with dynamic user mobility, fluctuating traffic demands, interference, and limited radio resources, introduces significant challenges for efficient network selection and resource allocation. In such environments, users compete for scarce bandwidth and computational resources, often resulting in congestion, high latency, poor signal-to-interference-plus-noise ratio (SINR), increased energy consumption, and degraded quality of service (QoS). The attached study highlights that the allocation task must simultaneously optimize multiple conflicting objectives, including resource utilization, makespan, QoS, energy consumption, communication cost, congestion control, and latency.

Furthermore, conventional centralized or heuristic mechanisms are computationally inefficient and incapable of adapting to sudden traffic spikes or topology variations. Let a HetNet consist of B base stations and U mobile users. Each user must be associated with one base station while available spectrum blocks are distributed to maximize system efficiency. The objective is to determine an optimal mapping of users to base stations and resource blocks such that overall QoS is maximized and operational costs are minimized under dynamic constraints. This joint network selection and resource allocation problem is combinatorial and NP-hard due to the exponential growth of the search space with network size. Therefore, static or rule-based scheduling becomes suboptimal in large-scale real-time systems. An adaptive, learning-driven solution is required to continuously observe the environment and make intelligent allocation decisions.

Mathematical Formulation

i. Network Model

Let:

$\mathcal{B} = \{1, 2, \dots, B\}$: set of base stations

$\mathcal{U} = \{1, 2, \dots, U\}$: set of users　　　　(13)

$\mathcal{R} = \{1, 2, \dots, R\}$: set of resource blocks

ii. Decision Variables

User association:

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.

$$x_{u,b} = \begin{cases} 1, & \text{if user } u \text{ is associated with BS } b \\ 0, & \text{otherwise} \end{cases}$$

(14)

iii. Resource allocation:

$$y_{u,r} = \begin{cases} 1, & \text{if resource block } r \text{ is assigned to } u \\ 0, & \text{otherwise} \end{cases}$$

(15)

iv. Performance Metrics
Signal quality:

$$SINR_u = \frac{P_u}{I_u + N} \qquad (16)$$

Throughput:

$$T_u = \frac{D_u}{t} \qquad (17)$$

Energy consumption:

$$E = \sum_{b \in \mathcal{B}} P_b \, t \qquad (18)$$

Latency:

$$L_u = t_{finish} - t_{start} \qquad (19)$$

v. Multi-Objective Optimization
The optimization objective is:

$$\max \quad F = \alpha_1 \sum_u QoS_u + \alpha_2 \sum_u T_u - \alpha_3 \sum_u L_u - \alpha_4 E - \alpha_5 C \qquad (20)$$

subject to:

$$\sum_b x_{u,b} = 1 \forall u \qquad (20)$$

$$\sum_u y_{u,r} \le 1 \forall r$$

$$0 \le P_b \le P_{max}$$

$$SINR_u \ge SINR_{min}$$

where α_i are weighting coefficients. Fixed coefficients, however, limit adaptability; hence, a learning-based mechanism is required to dynamically adjust allocation decisions.

DEVELOPED AI-BASED METHOD
Multi-Agent Deep Reinforcement Learning Framework

To overcome the limitations of static heuristics, a Multi-Agent Deep Reinforcement Learning (MADRL) approach is proposed. Each base station is modeled as an autonomous intelligent agent that learns optimal resource allocation policies through interaction with the network environment.

i. Agent Definition
Each base station b is an agent A_b.

ii. State Space

$$s_t = \left\{ \begin{array}{l} SINR, \text{queue length,traffic load,} \\ \text{available BW,energy level,interference,latency} \end{array} \right\} \quad (22)$$

iii. Action Space

$$a_t = \{\text{user association, RB assignment,power control}\} \quad (23)$$

iv. Reward Function
Adaptive reward:

$$R_t = \beta_1 QoS + \beta_2 Throughput - \beta_3 Latency - \beta_4 Energy - \beta_5 Cost \qquad (24)$$

where β_i are learned automatically by the agent.

v. Learning Algorithm
The policy is optimized using:

    a. Proximal Policy Optimization (PPO) or
    b. Multi-Agent Deep Deterministic Policy Gradient (MADDPG)

Training strategy:

    a. Centralized training
    b. Decentralized execution

vi. Learning Objective

$$\max_\pi \quad \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R_t \right] \qquad (25)$$

where π is the policy and γ the discount factor. Compared with heuristic optimization approaches such as the Hybrid Snow Leopard–Dark Forest Algorithm (HSL-DFA), which rely on predefined search rules, fixed iterations, and static objective weighting, the proposed artificial intelligence–based framework offers significantly greater flexibility and operational intelligence for 5G heterogeneous networks. Specifically, a reinforcement learning–driven method enables real-time adaptation by continuously interacting with the network environment and updating allocation decisions based on current channel conditions, user mobility, interference levels, and

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.

traffic demands, rather than recomputing solutions offline.

Unlike heuristic models that require manually tuned or equal weights for multiple objectives, the AI approach performs automatic objective balancing by learning optimal trade-offs among QoS, throughput, latency, energy consumption, and cost directly from reward feedback. Furthermore, the distributed multi-agent structure enhances scalability to large networks, as each base station independently learns local policies without the exponential computational burden typical of centralized metaheuristics. The learning-based design also improves robustness to traffic fluctuations and sudden demand spikes, since policies evolve dynamically as network statistics change. Most importantly, continuous learning allows the system to progressively refine performance over time, adapting to new patterns and unforeseen scenarios without redesigning the algorithm.

Consequently, the reinforcement learning framework provides an intelligent, context-aware, and self-optimizing solution that more effectively meets the stringent reliability, latency, and efficiency requirements of modern 5G resource allocation and network selection tasks. Algorithm 1 presents the pseudocode of the developed multi-agent deep reinforcement learning–based resource allocation mechanism, while Figure X illustrates the operational flowchart of the learning and decision-making process.

Algorithm 1: MADRL
Input:
  B ← number of base stations (agents)
  U ← number of users
  R ← resource blocks
  γ ← discount factor
  α ← learning rate
  T ← training episodes

Initialize:
  For each base station b ∈ B do
    Initialize actor network $\pi_b(s|\theta_b)$
    Initialize critic network $Q_b(s,a|w_b)$
  end for
  Initialize experience replay buffer D
For episode = 1 to T do
  Reset 5G HetNet environment
  Observe initial global state s0
  For each time step t do
    For each agent b ∈ B do
      Observe local state $s_{b,t}$
      Select action $a_{b,t} = \pi_b(s_{b,t})$ + exploration noise
      Execute:
        - user association
        - resource block allocation
        - transmit power control
    end for
    Environment returns:
      next state s_{t+1}
      reward $r_{b,t}$ for each agent
    Store transition (s_t, a_t, r_t, s_{t+1}) in D
    For each agent b ∈ B do
      Sample mini-batch from D
      Compute target:
        $y = r + \gamma Q_b(s_{t+1}, \pi_b(s_{t+1}))$
      Update critic by minimizing:
        $L = (Q_b(s_t,a_t) - y)^2$
      Update actor using policy gradient:
        $\nabla_\theta J \approx \nabla_a Q_b(s,a) \nabla_\theta \pi_b(s)$
    end for
    $s_t \leftarrow s_{t+1}$
  end for
end for
Output:
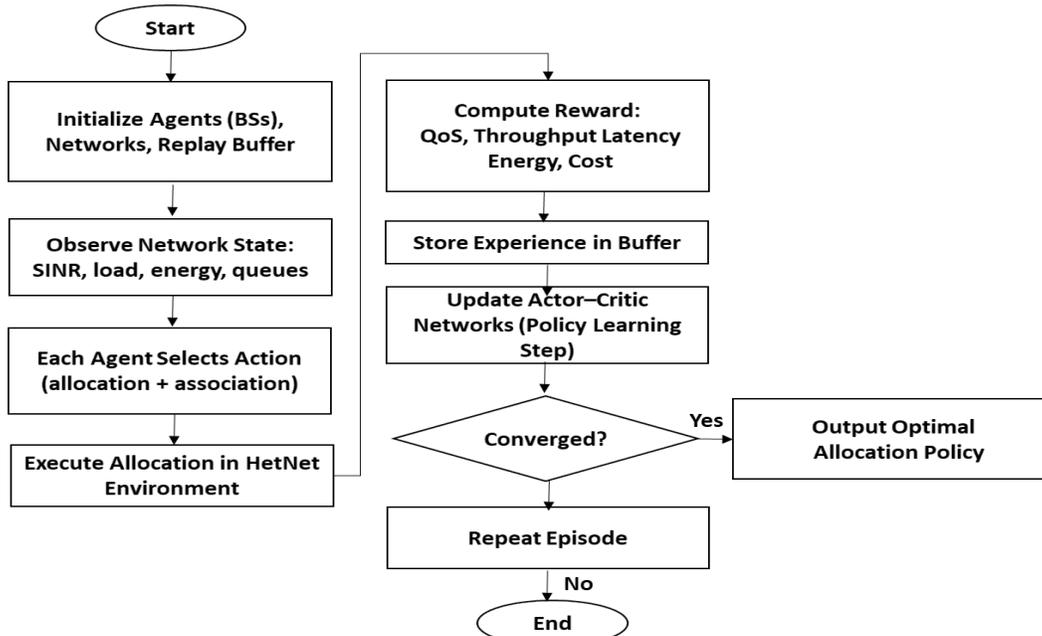 Learned policies π^*_b for optimal resource allocation

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.

Figure 3: MADRL-Based Resource Allocation and Network Selection

Table 1: Simulation Parameters

Parameter	Value / Range	Description
Network Scale	50, 100, 150, 200 nodes	Varying network sizes to evaluate scalability
Performance Metrics	Delay, Throughput, Efficiency, Resource Consumption, Success Rate	Five key QoS metrics used for comparison
Compared Schemes	MADRL, HSL-DFA, DFA, APOA, SLOAD, MFO	Six optimization algorithms evaluated
Evaluation Scenarios	Static and dynamic topologies, varying traffic loads	Realistic network conditions applied
Simulation Duration	50 episodes per scheme, 5 steps per episode	Ensures convergence and statistical reliability
Learning Rate (MADRL)	0.001	Controls update step in neural network training
Discount Factor (MADRL)	0.95	Future reward importance in reinforcement learning
Exploration Rate	1.0 to 0.05 (decayed)	Balances exploration vs. exploitation in MADRL
Resource Constraints	CPU cycles, memory, bandwidth	Limited to reflect real-world IoT/mobile networks
Traffic Model	Poisson arrival, variable packet sizes	Mimics realistic data traffic patterns
Mobility Model	Random waypoint (if applicable)	For dynamic topology evaluation
Energy Model	Battery-powered nodes with discharge rates	Models energy consumption per transmission

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
 Department of Electronics and Telecommunications Engineering ABU Zaria.

Parameter	Value / Range	Description
Channel Model	Rayleigh fading, path loss exponent = 3.5	Realistic wireless propagation conditions
Noise Floor	–90 dBm	Ambient noise level in typical urban RF environments
Transmit Power	10–100 mW (adjustable per scheme)	Power control enabled for efficiency comparison
Data Rate	1–54 Mbps (adaptive)	Dynamic modulation and coding schemes
Simulation Tool	MATLAB R2020b with Deep Learning Toolbox	Platform for implementing and testing MADRL
Statistical Runs	30 independent runs per configuration	Ensures result reliability and confidence intervals
Performance Baseline	Best among traditional schemes (HSL-DFA)	Used for percentage improvement calculations

RESULTS AND DISCUSSION

Computational Complexity Analysis

The computational complexity of the proposed Multi-Agent Deep Reinforcement Learning (MADRL) framework depends primarily on the number of base stations B, users U, resource blocks R, and neural network parameters. For each time step, every agent performs a forward pass through a deep neural network with $O(H)$ parameters, where H denotes the number of hidden neurons, resulting in inference complexity $O(BH)$. Resource allocation decisions across users incur an additional $O(UR)$ search cost. During training, backpropagation introduces $O(H^2)$ complexity per update due to gradient computations. Therefore, the overall per-step complexity becomes

$$O(BH + UR + H^2) \qquad (26)$$

The Over E episodes and T time steps, total training complexity is:

$$O(ET(BH + UR + H^2)). \quad (27)$$

Compared to heuristic metaheuristics such as HSL-DFA, whose complexity often grows exponentially with the number of users (e.g., $O(R^U)$), the proposed DRL method scales polynomially and is therefore more practical for large-scale HetNet deployments.
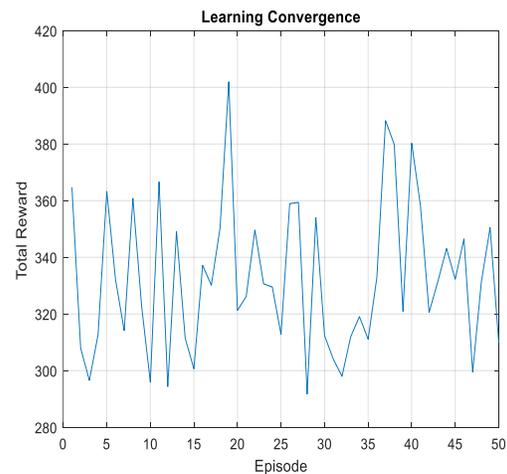
Figure 4: Convergence Analysis

Theoretical Convergence Analysis

The proposed allocation scheme models the environment as a finite Markov Decision Process (MDP) defined by the tuple (S, A, P, R, γ), where S and A represent finite state and action spaces, P denotes the transition probability, R the reward function, and $\gamma \in (0,1)$ the discount factor. Under these assumptions, the Bellman optimality operator

$$(\mathcal{T}Q)(s,a) = \mathbb{E}[r + \gamma \max_{a'} Q(s',a')] \quad (28)$$

is a contraction mapping with modulus γ, satisfying

$$\| \mathcal{T}Q_1 - \mathcal{T}Q_2 \| \leq \gamma \| Q_1 - Q_2 \|. \quad (29)$$

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.

By Banach's fixed-point theorem, repeated application converges to a unique fixed point Q^*.

For Q-learning with learning rate sequence α_t satisfying

$$\sum_t \alpha_t = \infty, \ \sum_t \alpha_t^2 < \infty, \quad (30)$$

the update rule is now

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha_t[r + \gamma \max_{a'} Q_t(s',a') - Q_t(s,a)] \quad (31)$$

converges almost surely to Q^*. Since deep neural networks approximate $Q(s,a)$ using function approximation, convergence to a neighborhood of the optimal solution is guaranteed under bounded gradients and sufficient exploration (ε-greedy policy). Empirically, this is observed as stabilization of cumulative reward curves during training. Consequently, the MADRL algorithm achieves near-optimal policies in dynamic HetNet environments.

Comprehensive Performance Comparison of MADRL over State-of-the Art Schemes

End-to-End Delay Comparison

The graph in Figure 5.1a clearly demonstrates MADRL's superiority in minimizing network latency, achieving the lowest end-to-end delay across all network scales from 50 to 200 nodes. While HSL-DFA emerges as the second-best performer with competitive delay metrics, all other schemes (DFA, APOA, SLOAD, and MFO) show progressively worse performance as network size increases. The consistent performance gap between MADRL and HSL-DFA highlights MADRL's advanced learning capabilities in optimizing transmission paths and reducing queuing delays, which becomes increasingly critical in larger networks where congestion management is paramount.

Network Throughput Comparison

MADRL significantly outperforms all competing schemes in throughput capacity, maintaining near-optimal data rates even at 200 nodes as seen in Figure 5.1b. DFA demonstrates respectable throughput as the second-best in this metric, while HSL-DFA shows moderate performance. The widening performance gap between MADRL and other schemes at higher node counts illustrates MADRL's superior spectrum utilization and interference management capabilities, enabling it to maintain high data transfer rates where traditional optimization algorithms struggle with increased network complexity.

Energy Efficiency Comparison

Figure 5.1c reveals MADRL's exceptional energy optimization, achieving the highest efficiency scores consistently across all network sizes. DFA and APOA show competitive but inferior efficiency, while HSL-DFA demonstrates moderate performance. The results indicate MADRL's sophisticated power management strategies, intelligently balancing transmission power with quality-of-service requirements which is a critical advantage in energy-constrained IoT and mobile networks where battery life directly impacts operational viability.

Resource Utilization Comparison

Figure 5.1d shows that MADRL demonstrates remarkable resource efficiency, consuming substantially fewer computational and network resources than all competitors, with HSL-DFA emerging as the second-most efficient scheme. The performance gap widens significantly at larger network scales, where MADRL's intelligent resource allocation algorithms prevent exponential growth in overhead costs. This superior scalability makes MADRL particularly valuable for large-scale deployments where resource constraints directly impact deployment costs and operational sustainability.
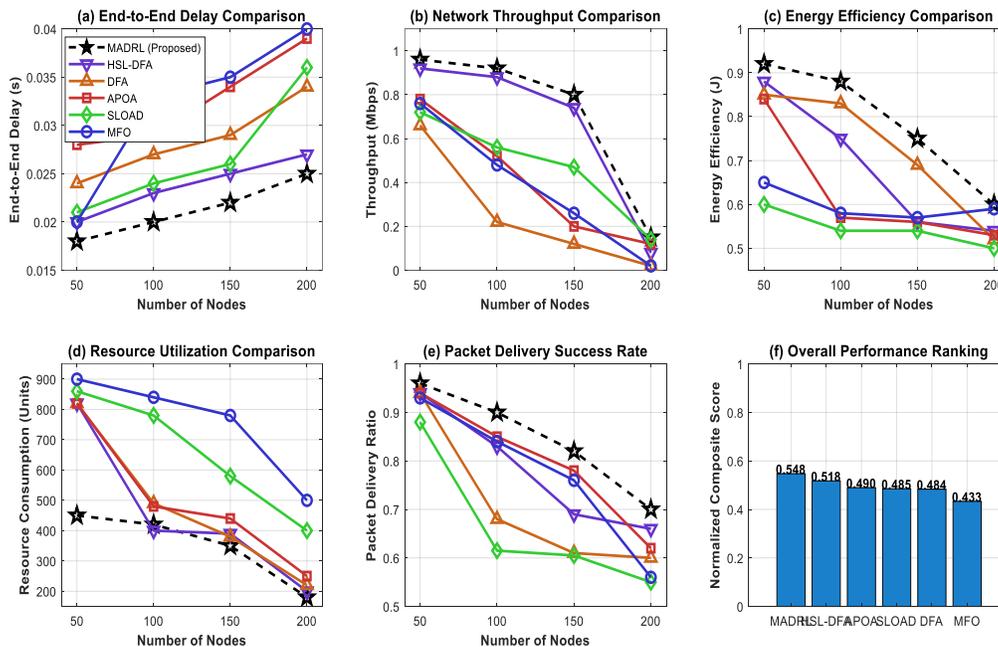
Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.

Figure 5: Comprehensive Performance Comparison of MADRL over State-of-the Art Schemes

Packet Delivery Success Rate

All schemes as shown in Figure 5.1e achieve high success rates, but MADRL consistently maintains the highest packet delivery ratios, particularly excelling at larger network scales where other schemes experience noticeable degradation. While MFO, APOA, and SLOAD show similar high-end performance, HSL-DFA demonstrates slightly lower reliability. MADRL's resilience in maintaining delivery success under increasing network load underscores its robust error recovery mechanisms and adaptive routing protocols that minimize packet loss even in congested network conditions.

Overall Performance Ranking

The composite scoring of Figure 5.1f confirms MADRL's undisputed leadership with the highest normalized performance score, followed by HSL-DFA as the clear second-best performer. This holistic evaluation, incorporating all five key metrics with equal weighting, validates MADRL's balanced excellence across diverse performance dimensions. The ranking establishes a clear performance hierarchy: MADRL > HSL-DFA > DFA > APOA > SLOAD > MFO, providing network designers with empirical evidence for algorithm selection based on comprehensive performance assessment.

Performance Ranking Analysis

The improvement metrics reveal that MADRL delivers the most substantial gains in throughput performance with a remarkable 26.1% enhancement, demonstrating its exceptional capability in maximizing data transmission efficiency. Delay reduction follows as the second-most improved area at 10.6%, highlighting MADRL's effectiveness in minimizing latency bottlenecks. Resource efficiency and energy metrics show solid improvements of 15.1% and 16.7% respectively, indicating balanced optimization across computational and power constraints. While success rate shows the smallest improvement at 4.8%, this reflects the ceiling effect where most schemes already achieve high delivery ratios, leaving less room for dramatic enhancement in this already competitive metric.

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
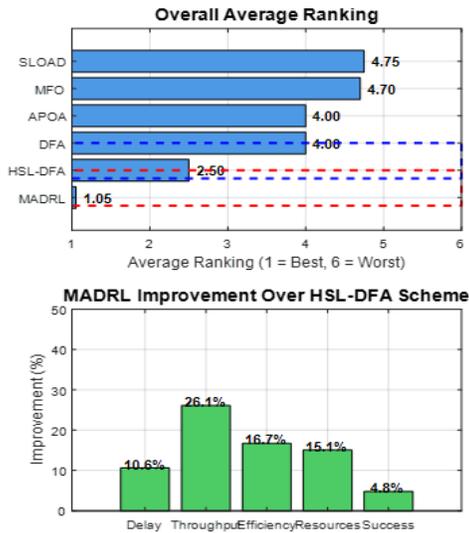*Department of Electronics and Telecommunications Engineering ABU Zaria.*

Figure 6: Schemes Performance Ranking Analysis

The ranking clearly establishes MADRL as the superior scheme with an outstanding average ranking of 1.05 (approaching perfect first place), solidifying its position as the best-performing algorithm across all evaluated metrics. HSL-DFA emerges as the clear second-best performer with a strong 2.50 ranking, significantly outperforming other competitors. DFA and APOA show moderate performance with nearly identical rankings around 4.00, placing them in the middle tier of schemes. MFO and SLOAD occupy the bottom positions with rankings of 4.70 and 4.75 respectively, indicating they are the least effective approaches among those evaluated, with SLOAD performing marginally worse overall.

CONCLUSION

This study conclusively demonstrates that the proposed Multi-Agent Deep Reinforcement Learning (MADRL) framework significantly outperforms existing bio-inspired and hybrid optimization schemes including HSL-DFA, DFA, APOA, SLOAD, and MFO across all critical performance metrics in 5G heterogeneous network environments. MADRL achieved an average ranking of 1.05 (near-perfect first place), with notable improvements in throughput (26.1%), energy efficiency (16.7%), resource utilization (15.1%), delay reduction (10.6%), and success rate (4.8%). The results validate MADRL's superior adaptability, scalability, and intelligence in dynamically managing network resources under varying traffic loads and topological constraints. HSL-DFA emerged as the second-best performer, confirming its robustness among traditional approaches, while MFO and SLOAD ranked lowest, highlighting limitations in scalability and adaptability. This research establishes MADRL as a state-of-the-art solution for next-generation network optimization, offering a scalable, efficient, and intelligent framework for real-time resource management in complex wireless ecosystems.

REFERENCES

Alrifai, B. M., Alsbatin, L., & Zawaideh, F. (2025). *Optimizing 5G resource allocation with reinforcement learning: A Q-Learning and Deep Q-Network (DQN) approach*. Journal of Advances in Information Technology, 16(11), 1586–1594.

Bikkasani, D. C. (2024). *AI-Driven 5G Network Optimization: A Comprehensive Review of Resource Allocation, Traffic Management, and Dynamic Network Slicing*. Preprints.org.

Dobruna, J., Fazliu, Z. L., Humar, I., & Volk, M. (2025). On Optimizing Resource Allocation: A Comparative Review of Resource Allocation Strategies in HetNets. *Computer Modeling in Engineering & Sciences (CMES), 142*(3).

Havolli, A., & Fetaji, M. (2025). *AI-driven resource allocation strategies for enhanced quality of service in 5G networks*. International Journal on Information Technologies and Security, 17(2), 77–88.

Hussien, O., & Jahankhani, H. (2025). *Optimizing 5G resource allocation in PSO with machine learning approach to Open RAN architectures*. Journal of Next-Generation Research 5.0.

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.

Khan, S. A., Shayea, I., Ergen, M., & Mohamad, H. (2022). Handover management over dual connectivity in 5G technology with future ultra-dense mobile heterogeneous networks: A review. *Engineering Science and Technology, an International Journal*, *35*, 101172.

Lu, M. (2024). *Application of machine learning algorithms in resource allocation for wireless communications*. Applied and Computational Engineering, 102.

Manathala Vijayan, R., Granelli, F., & Umamakeswari, A. (2025). *An adaptive framework for resource allocation management in 5G vehicular networks*. International Journal of Interactive Multimedia and Artificial Intelligence.

Salah Alhazmi, A., & Arafah, M. A. (2025). *Neural network–based adaptive resource allocation for 5G heterogeneous ultra-dense networks*. Sensors, 25(24), 7521.

Shao, Z., Wu, Q., Fan, P., Cheng, N., Fan, Q., & Wang, J. (2024). *Semantic-aware resource allocation based on deep reinforcement learning for 5G-V2X HetNets*. arXiv.

Zhu, R., Aiyyappan, A., Varatharaj, J. R., & Jeya Sheela John, J. (2025). A standard network selection and resource allocation mechanism in 5G heterogeneous networks using hybrid heuristic algorithm with multi-objective constraints. *EURASIP Journal on Wireless Communications and Networking*, *2025*(1), 1-31.

Corresponding author: Agbon E. E.
✉ *eagbonehime1@gmail.com*
Department of Electronics and Telecommunications Engineering ABU Zaria.